

RODRIGO AQUINO DE CARVALHO

**PERSPECTIVAS NA WEB SEMÂNTICA PARA A
CIÊNCIA DA INFORMAÇÃO**

**CAMPINAS
2009**

RODRIGO AQUINO DE CARVALHO

**PERSPECTIVAS NA WEB SEMÂNTICA PARA A
CIÊNCIA DA INFORMAÇÃO**

Dissertação apresentação ao Programa de Pós-Graduação em Ciência da Informação da Pontifícia Universidade Católica de Campinas, como requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

Orientadora: Prof^ª Dr^ª Mariângela Pisoni Zanaga.

Área de concentração: Administração da Informação.

Linha de Pesquisa: Gestão da Informação.

**CAMPINAS
2009**

AUTORIZO A **REPRODUÇÃO PARCIAL OU TOTAL** DESTE TRABALHO, POR QUALQUER MEIO, CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE, SEJA CITADA A FONTE.

AGRADEÇO A **COMUNICAÇÃO DE QUALQUER EQUÍVOCO OU OMISSÃO** PRESENTE NO TRABALHO (Contato: rodrigo@yahoo.com.br).

Ficha Catalográfica Elaborada pelo Sistema de Bibliotecas e Informação – SBI – PUC-Campinas

t020
C331p

Carvalho, Rodrigo Aquino de
Perspectivas na Web Semântica para Ciência da Informação. – Campinas : PUC-Campinas, 2009.
186f.

Orientadora: Prof^ª Dr^a Mariângela Pisoni Zanaga
Dissertação (Mestrado) – Pontifícia Universidade Católica de Campinas, Centro de Ciências Humanas e Sociais Aplicadas, Programa de Pós-Graduação em Ciência da Informação.
Inclui apêndices e bibliografia.

1. Ciência da Informação 2. Web Semântica. 3. Recuperação da informação na Web. 4 Representação da informação na Web. 5. Representação do conhecimento. 6. Ciência da Informação – Teses e dissertações I. Zanaga, Mariângela Pisoni II. Pontifícia Universidade Católica de Campinas. Centro de Ciências Humanas e Sociais Aplicadas. Programa de Pós-Graduação em Ciência da Informação III. Título.

22.ed. CDD – t020

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE CAMPINAS
CENTRO DE CIÊNCIAS HUMANAS E SOCIAIS APLICADAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Autor (a): CARVALHO, Rodrigo Aquino de .

Título: "PERSPECTIVAS NA WEB SEMÂNTICA PARA A CIÊNCIA DA INFORMAÇÃO"

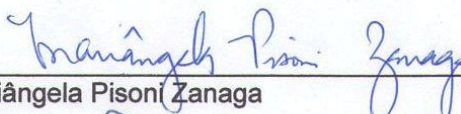
Orientadora: Profa. Dra. Mariângela Pisoni Zanaga

Dissertação de Mestrado em Ciência da Informação

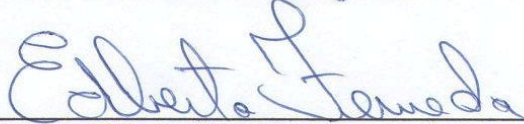
Este exemplar corresponde à redação final da Dissertação de Mestrado em Ciência da Informação da PUC-Campinas, e aprovada pela Banca Examinadora.

Data: 27/02/2009.

BANCA EXAMINADORA



Profa. Dra. Mariângela Pisoni Zanaga



Prof. Dr. Edberto Farneda



Prof. Dr. José Oscar Fontanini de Carvalho

Dedico

Aos meus pais, Paulo e Izaura

À toda a minha família

À minha noiva Priscilla

AGRADECIMENTOS

Agradeço a DEUS, por me dar essa oportunidade de realizar algo... Obrigado PAI!

À toda minha família, principalmente meus pais.

À professora Mariângela, pela orientação e paciência, mas principalmente pela confiança e respeito. Obrigado.

À professora Vera Beraquet, que me colocou no caminho da pesquisa e me ensinou muitas outras coisas.

Aos professores participantes da banca, José Oscar Carvalho e Edberto Fernalda, pela atenção em um momento muito importante.

Aos demais professores: Maria de Fátima Tálamo, Rogério Bazi e Leonardo Garcia.

Aos queridos amigos: Alexander, Claudete, Fabiana, Graziela, João, Jonatas, Marcos, Renatha James, Rita e Sonia (em ordem alfabética).

Aos funcionários da PUC-Campinas que auxiliaram muito na construção desse momento, em especial aos da secretaria acadêmica do programa e aos da biblioteca do CCSA.

À CAPES, pela concessão da bolsa de estudos.

À PUC-Campinas, por ceder seu espaço e prestígio para a realização de importante etapa da minha vida.

À Priscilla Christine, minha amada e querida noiva, por entender as ausências, pela oração e apoio. Obrigado.

A palavra progresso não tem nenhum sentido enquanto ainda existirem crianças infelizes.

Albert Einstein

RESUMO

CARVALHO, R. A. de. **Perspectivas na Web Semântica para a Ciência da Informação**. Campinas, 2008. 186f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Pontifícia Universidade Católica de Campinas. 2008.

A Web se tornou o principal veículo de armazenamento e disseminação da informação, mas há a necessidade de que a informação nesse ambiente seja tratada de forma eficiente. O objetivo da pesquisa é buscar perspectivas na Web Semântica para a Ciência da Informação, através do mapeamento na literatura das propostas básicas dessa tecnologia e do desenvolvimento já alcançado e também da análise das teses e dissertações sobre o tema apresentado aos Programas de Pós-graduação em Ciência da Informação no Brasil. O método utilizado foi o dialético, com abordagem exploratória e descritiva, o procedimento técnico foi o bibliográfico e as técnicas para a estruturação dos dados foram a bibliometria e a análise de conteúdo. Foram estudados os seguintes movimentos e relações: - do surgimento da Web até o estado atual da Web Semântica (mudança qualitativa); - Web atual e Web Semântica (movimento, mudança qualitativa); - Ciência da Informação (através da revisão de literatura) e Web semântica (relações); e - Ciência da Informação (em teses e dissertações do Brasil) e Web Semântica (relações). O movimento da Web atual à Web Semântica foi significativo, com objetivos voltados a possibilitar inferências complexas por parte das máquinas, mas deixando a simplicidade da arquitetura da Web atual, a Web Semântica se desenvolveu pouco, não teve movimento qualitativo significativo. Quanto à CI, sua arquitetura sustenta uma relação. A literatura da CI, no que diz respeito à recuperação e representação da informação e representação do conhecimento, possui recursos que podem colaborar para o desenvolvimento da Web Semântica (como os tesauros, os metadados etc.). Outro ponto que favorece essa aproximação é a relação interdisciplinar da área com a Ciência da Computação, com a Inteligência Artificial e com a Linguística (Terminologia). Mas existe a necessidade de que o conhecimento gerado na CI, sobre a Web Semântica ultrapasse os limites da própria área.

PALAVRAS-CHAVE: Ciência da informação. Web Semântica. Recuperação da Informação na Web. Representação da Informação na Web. Ciência da Informação – Teses e dissertações.

ABSTRACT

CARVALHO, R. A. de. **Perspectives in the Semantic Web to the Information Science**. Campinas, 2008. 186f. Dissertation (Master in Information Science) – Programa de Pós-Graduação em Ciência da Informação, Pontifícia Universidade Católica de Campinas. 2008.

The Web became the main vehicle of the information storage and dissemination, but the information in this environment must be organized by efficient manner. The research objective was identify perspectives in the Semantic Web to the Information Science, through the literature mapping of basic proposals of this technology and of its development and through the analysis of the theses and dissertations about the subject presented to the Graduate Programs of Information Science in Brazil. Dialectical method with exploratory and descriptive approach and bibliographical technical procedures allied with bibliometrics and content analysis were applied. The following movements and relations were studied: - the origin of the Web to the actual status of the Semantic Web (qualitative change); - the actual Web and the Semantic Web (movement, qualitative change); - Information Science (through the literature review) and Semantic Web (relations); and – Information Science (in theses and dissertations of the Brazil) and Semantic Web (relations). The movement of the current Web to the Semantic Web was significant, with objectives dedicated to make possible complex inferences about machines. Leaving the simplicity of the architecture of the current Web, the Semantic Web grew little, it didn't have significant qualitative movement. With relationship to Information Science, your architecture sustains a relationship. The field of Information Science, in what concerns to the literature of retrieval and representation of the information and representation of the knowledge presents resources that can collaborate for the development of the Semantic Web (as thesaurus, metadata etc.). Other point that favours that approach is the interdisciplinary relationship of the field with the Computational Science, with the Artificial Intelligence and with the Linguistics (Terminology). But it exists the need that the knowledge on the generated knowledge Information Science on the Semantic Web surpasses the limits of the proper field.

KEY-WORDS: Information Science. Semantic Web. Information retrieval in Web. Information representation in Web. Information Science – Theses and dissertations.

LISTA DE ABREVIATURAS E SIGLAS

AC	Análise de conteúdo
AD	Análise do discurso
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior
CC	Ciência da Computação
CI	Ciência da Informação
DC	Dublin Core
IA	Inteligência Artificial
OI	Organização da informação
PUC-Campinas	Pontifícia Universidade Católica de Campinas
RC	Representação do conhecimento
RI	Recuperação da Informação
TICs	Tecnologias de Informação e Comunicação
UFBA	Universidade Federal da Bahia
UFF	Universidade Federal Fluminense
UFMG	Universidade Federal de Minas Gerais
UFPB	Universidade Federal da Paraíba
UFSC	Universidade Federal de Santa Catarina
UnB	Universidade de Brasília
UNESP	Universidade Estadual Paulista
USP	Universidade de São Paulo
W3C	World Wide Web Consortium
WS	Web Semântica

LISTA DE QUADROS

Quadro 1.	Termos-chave para pesquisa bibliográfica	25
Quadro 2.	Especificidades da análise de conteúdo	28
Quadro 3.	Categorias para a análise de conteúdo	32
Quadro 4.	Etapas da análise de conteúdo utilizadas no trabalho	33
Quadro 5.	Exemplo da extração dos temas de um parágrafo.....	34
Quadro 6.	Relações Interdisciplinares da CI na proposta de Lena Vânia Ribeiro Pinheiro	59
Quadro 7.	Relações Interdisciplinares da CI na proposta de Saracevic.....	60
Quadro 8.	Relações Interdisciplinares propostas na presente pesquisa	60
Quadro 9.	Classes principais da Classificação Decimal de Dewey (CDD).....	80
Quadro 10.	Divisão da classe ‘Ciências Puras’ (500) da CDD	80
Quadro 11.	Classes principais da Classificação Decimal Universal (CDU)	81
Quadro 12.	Divisão da classe ‘Ciência Aplicada. Medicina. Tecnologia’ (6) da CDU	81
Quadro 13.	Tipos de indexação	84
Quadro 14.	Estrutura básica de metadados do Formato MARC.....	86
Quadro 15.	Sumário do ‘Docutes: tesouro de Ciencias de la documentación’	90
Quadro 16.	Classes principais do ‘Docutes: tesouro de Ciencias de la documentación’	90
Quadro 17.	Parte do índice hierárquico do ‘Docutes: tesouro de Ciencias de la Documentación’	93
Quadro 18.	Diferenças entre taxonomias, sistemas de classificação e tesouros	94
Quadro 19.	Diferenças entre taxonomias, sistemas de classificação, tesouros e ontologias	97
Quadro 20.	A entropia como fundamento da Cibernética	103
Quadro 21.	Comparação do computador com o cérebro humano	104
Quadro 22.	Algumas diferenças entre a Web social e a Web Semântica	131
Quadro 23.	Distribuição dos programas de pós-graduação em Ciência da Informação no Brasil e das teses e dissertações recuperadas para a pesquisa	133
Quadro 24.	Teses da área de CI relacionadas ao tema Web Semântica	135
Quadro 25.	Dissertações da área de CI relacionadas ao tema Web Semântica	136
Quadro 26.	Palavras-chaves dos trabalhos que se relacionam com o termo ‘Web Semântica’	141
Quadro 27.	Distribuição das palavras-chave das dissertações com autores com formação em nível de graduação na área de Biblioteconomia	141
Quadro 28.	Distribuição das palavras-chave dos trabalhos de autores com formação em nível de graduação na área de informática.....	142

LISTA DE FIGURAS

Figura 1.	Relação do método dialético com as técnicas de estruturação e tratamento dos dados.....	20
Figura 2.	Percepções sobre a Ciência da Informação enquanto disciplina científica.....	44
Figura 3.	Hexágono cognitivo e seus seis campos constituintes.....	63
Figura 4.	Intersecção entre a Linguística e a Ciência da Informação	68
Figura 5.	Cenário de recuperação da informação.....	70
Figura 6.	Arquitetura básica de um sistema de recuperação da informação	71
Figura 7.	Uso de operadores lógicos na recuperação da informação	75
Figura 8.	Cenário de organização e representação do conhecimento	78
Figura 9.	Quatro visões sobre a tecnologia	101
Figura 10.	Principais fatos para a criação e o desenvolvimento da Internet e da Web antes da W3C.....	112
Figura 11.	Desenvolvimento da Web a partir da criação da W3C em 1994	113
Figura 12.	Estrutura tecnológica da World Wide Web	117
Figura 13.	Distribuição dos textos das duas primeiras dialéticas de análise.....	119
Figura 14.	Arquitetura da Web sintática em 1998	122
Figura 15.	Modelo geral de estrutura de descrição de recursos (RDF).....	124
Figura 16.	Exemplo de aplicação do RDF ao conjunto de Metadados DC.....	124
Figura 17.	Arquitetura da Web Semântica.....	127
Figura 18.	Proporção das teses e dissertações recuperadas.....	134
Figura 19.	Distribuição da teses e dissertações por ano de defesa.....	134
Figura 20.	Formação acadêmica dos autores da teses.....	137
Figura 21.	Formação acadêmica dos autores das dissertações.....	138
Figura 22.	Distribuição das palavras-chave das teses e dissertações	139
Figura 23.	Identificação de parte do trabalho onde ocorre o aparecimento do Termo ‘Web Semântica’	140
Figura 24.	Algumas convergências entre a Ciência da Informação e a Web.....	152
Figura 25.	Esquema de sugestão para a criação de níveis e setores para o desenvolvimento da WS.....	153

LISTA DE TABELAS

Tabela 1.	Distribuição mundial de usuários de Internet de 2000 a 2003 (em milhões).....	110
Tabela 2.	Uso da Internet no mundo em 2005 (percentuais)	110

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Objetivos	19
1.2 Procedimentos metodológicos.....	20
1.2.1 Método dialético	21
1.2.2 Abordagem metodológica exploratória e descritiva	24
1.2.3 Pesquisa bibliográfica.....	24
1.2.4 Técnicas metodológicas para estruturação e apresentação dos dados	25
1.2.4.1 Bibliometria	25
1.2.4.2 Análise de conteúdo.....	27
1.4 Apresentação dos capítulos	34
 2 CIÊNCIA DA INFORMAÇÃO	 36
2.1 O contexto científico da Ciência da Informação	36
2.2 Origem e períodos da Ciência da Informação	45
2.2.1 Biblioteconomia	46
2.2.2 Documentação	47
2.2.3 Tecnologias de informação e comunicação.....	49
2.3 Conceitos de Ciência da Informação	51
2.4 Objeto: a informação	53
2.5 Natureza interdisciplinar	56
2.5.1 Ciência da Computação.....	61
2.5.2 Ciências Cognitivas: Inteligência Artificial	63
2.5.3 Lingüística: Terminologia	65
 3 INFORMAÇÃO E CONHECIMENTO: recuperação, organização e representação	 69
3.1 Recuperação da informação	69
3.2 Representação da informação e representação do conhecimento.....	76
3.3 Classificação.....	79
3.4 Indexação.....	82
3.5 Metadados	85
3.6 Tesouros.....	88

3.7 Taxonomias	93
3.8 Ontologias	94
4 TECNOLOGIA E SOCIEDADE	98
4.1 Algumas reflexões sobre o impacto da tecnologia	98
4.2 A automação e os períodos da computação	103
4.3 Internet.....	108
4.4 World Wide Web.....	111
4.4.1 Outros precursores	117
5 PANORAMA DA WEB SEMÂNTICA.....	119
5.1 A Web sintática ou atual.....	120
5.2 A Web Semântica.....	123
6 A WEB SEMÂNTICA NA CIÊNCIA DA INFORMAÇÃO: um estudo a partir das teses e dissertações de programas de pós-graduação da área no Brasil.....	132
6.1 Dados gerais de produção das teses e dissertações.....	133
6.2 Formação acadêmica dos autores das teses e dissertações	137
6.3 Distribuição das palavras-chave	139
6.4 Análise individual dos trabalhos com a ocorrência do termo ‘Web Semântica’ no resumo	143
6.5 Análise individual dos trabalhos com a ocorrência do termo ‘Web Semântica’ no título ou nas palavras-chave	144
7 CONSIDERAÇÕES FINAIS.....	150
REFERÊNCIAS	157
APÊNDICES.....	166
Apêndice A. Quadros da análise de conteúdo dos textos selecionados para identificação das propostas e desenvolvimento da Web Semântica	166
Apêndice B. Teses e dissertações analisadas no trabalho	176
Apêndice C. Palavras-chave indicadas nas teses e dissertações analisadas	186

1 INTRODUÇÃO

Um estudo sobre a representação da informação e do conhecimento e da recuperação da informação (RI), por vezes remete à como a sociedade utilizou a linguagem na realização de suas atividades, pois a linguagem, sendo um “sistema de articulação da mente” (KERCKHOVE, 2003, p.7), materializa a compreensão que o homem tem do mundo, assim como suas questões e reflexões. As sociedades orais possuíam dificuldades técnicas para registrar sua memória, sendo assim, eram os próprios indivíduos os responsáveis por fazer com que sua história continuasse. Kerckhove (2003) denomina esse tipo de sociedade, de sociedade do contexto, afirmando que elas não estudavam o passado, mas “viviam o passado no presente”; sua memória estava registrada em estátuas, monumentos, canções, histórias e atuações, e devido a isso, há até hoje pouco conhecimento sobre elas.

Surge uma nova tecnologia, a escrita, e, portanto, as sociedades da escrita. Kerckhove (2003) afirma que o material impresso, que é a interface dominante nesse tipo de sociedade, separa o texto do contexto, não havendo mais a preocupação de perda do passado, pois o material impresso possibilita o arquivamento da memória dessa sociedade e impulsiona o futuro.

Essa forma de arquivamento da memória da sociedade escrita através de registros e seus respectivos estoques informacionais estende-se até nossos dias e pode ser exemplificado pelas bibliotecas, arquivos e centros de documentação. Os estoques informacionais passaram a apresentar um novo problema, conforme indicado por Bush (1945). Há muito registro arquivado e a forma da organização da informação nesses estoques, por meio de esquemas hierárquicos, já não parece suficiente para que o material seja recuperado, pois há muito material, gerando o que pode ser chamado de “*caos informacional*”.

Surgem outras tecnologias, as tecnologias de informação e comunicação (TICs), amparadas pela já existente eletricidade, que auxiliam na organização e na recuperação dos estoques informacionais da sociedade escrita, mas que cria uma nova forma de armazenar a memória e materializar a linguagem. Para essa fase, Chartier (2002a) cunha o termo textualidade eletrônica. Já Kerckhove (2003), a denomina de linguagem eletrônica, sendo essa o terceiro estágio da linguagem indicado pelo autor, os estágios anteriores são: linguagem oral e linguagem escrita. Nessa sociedade eletrônica, sugerida por Kerckhove (2003), a principal interface é a tela, é a sociedade do

hipertexto¹, que compartilha do texto e do contexto, além de permitir a conectividade; é a sociedade em rede.

O surgimento da forma eletrônica de difundir, armazenar e apresentar a informação trouxe consigo muitas transformações. Chartier (2002a) alerta para a oferta textual multiplicada por essa “forma” de informação, o que podemos considerar um paradoxo, se a mesma surgiu para resolver o já citado “*caos informacional*”. Como transformações que exigem cuidados em relação à informação eletrônica, podemos citar um possível monopólio lingüístico do idioma inglês, um provável enfraquecimento da idéia de autoria e responsabilidade de uma obra, mudança do que entendemos por leitor para co-autor, oferta superabundante de informação que vai além da capacidade de apropriação do leitor, assim conduzindo a humanidade para uma possível perda daquilo que nos define: memória (por exemplo: o excesso de informação), nome (por exemplo: enfraquecimento da idéia de autoria) e diferença (por exemplo: predomínio do idioma inglês), como bem indica Chartier (2002a, 2002b).

A informação eletrônica propõe ainda a idéia de uma biblioteca universal, deixando ao alcance de “*todos*” o conhecimento da humanidade, reunindo um “Congresso do Mundo”, como a idealizou Borges (1977)² citado por Chartier (2002a, p.12) em uma de suas prosas. Além dessa improvável utopia, há outras transformações com o surgimento da forma eletrônica da informação. No que diz respeito à Web, há a união do que Santaella (2001) denomina das três matrizes da linguagem e pensamento, ou seja, as três formas de como se “coisifica” o conhecimento da humanidade; são as matrizes: sonora, visual e verbal, que na Ciência da Informação (CI) entendemos como documentos: sonoro e audiovisual, imagético e textual.

Segundo Chartier (2002a), com a textualidade eletrônica há a alteração da ordem dos discursos, ou seja, por um único meio (o computador), temos diversos tipos de expressões: o livro (*e-book*), o artigo (periódicos eletrônicos), a música, o filme, a imagem, etc. Isto é, há o desaparecimento da materialidade que diferencia os discursos (CHARTIER, 2002a). No contexto da Web, Cabrera González (2007) discute essa questão, afirmando que esse meio de comunicação reúne em si todas as formas tradicionais de comunicação (TV, rádio e imprensa) e as suas próprias (portais, blogs, etc.). Assim, podemos afirmar que, além do desaparecimento da materialidade que diferencia os discursos, há também uma convergência de todos os meios, ou seja, temos na tela de computador: o livro, o artigo, a música, o filme, a imagem, a televisão, o rádio, o jornal, etc. Ou seja, os livros não

¹ Uma forma não-linear de apresentar e consultar informações, vinculando informações contidas em documentos, criando uma rede de associações complexas através de *links* (nós). (COSTA, 2003, p.254).

² BORGES, J. L. El Cogresso. In: _____. *El libro de arena*. Madrid: Alianza Editorial, 1977. p.27-54.

são mais apenas as “folhas encadernadas”, a música já não é apenas o registro no “disco compacto”, eles são também arquivos de computador.

Além da alteração da ordem dos discursos, Chartier (2002a) indica mais duas mudanças proporcionadas pela textualidade eletrônica. São elas: a ordem das razões, onde as “modalidades de argumentação” não acontecem mais de forma linear, mas sim de forma “aberta, clara e racional graças à multiplicação dos vínculos *hipertextuais*” (CHARTIER, 2002a, p.24), ou seja, à semelhança das idéias propostas por Bush (1945) para a organização da informação, por associação, como na mente humana, o que para Kerckhove (2003) trata-se do favorecimento das relações icônicas (imagens da memória ou *links*); a ordem das propriedades: a textualidade eletrônica permite que o leitor interfira no texto, mudando a concepção de propriedade literária e o *copyright*, assim como as características do texto (CHARTIER, 2002a).

A informação eletrônica se apresenta, portanto, e exige a redefinição de diversas categorias: a jurídica (direitos autorais); a estética; a administrativa (depósito legal), a biblioteconômica (catalogação, classificação ou descrição bibliográfica) que parecem ainda estar no paradigma textual do “papel” (CHARTIER, 2002b, p.117), isto é, da sociedade escrita (KERCKHOVE, 2003).

Mas é certo, como indica Cabrera González (2007), que este momento, exemplificado principalmente pela Web, com a Web 2.0 ou Web social, está sendo caracterizado pela mudança de posição no processo comunicativo do receptor, que não mais se mostra passivo, mas interfere diretamente nos conteúdos. A autora citada apresenta tal mudança como uma vantagem. Já Chartier (2002a) considera que a autoria e a responsabilidade sofrem mudanças nesse contexto que devam ser repensadas, para que não tragam prejuízos.

Chartier (2002b) alerta que a textualidade eletrônica vai encontrar indivíduos com a mesma capacidade cognitiva da sociedade escrita, ou seja, a textualidade eletrônica não alterou a capacidade de pensamento dos indivíduos, apesar de exercer alguma influência sobre ela; e que a facilidade de inserção de conteúdos na web pode afetar a sua qualidade.

A Internet e principalmente a World Wide Web³ são os veículos que realmente concretizam a idéia de textualidade eletrônica, pois a questão da informação toma um impulso impressionante no que diz respeito ao alcance global. Essas e outras TICs que ignoraram espaço e tempo na disseminação e na RI, tornam possível que todos sejam autores de conteúdos, ocupem lugar de destaque em e frente a centros tradicionais de informação, como as bibliotecas, arquivos, museus, centros de

³ Em uma tradução literal, “Teia de Alcance Mundial”.

documentação e outros. Alteraram, como afirma Lévy (1998, p.104), se referindo ao computador, “a experiência de pensamento, que era uma bricolagem artesanal, em um empreendimento sistemático em grande escala”, o que para Kerckhove (2003) é denominado de cognição compartilhada.

Essa nova dinâmica da informação evidenciada no espaço Web fez surgir diversos termos, como: hipertexto, ciberespaço⁴ e cibercultura⁵, indicando a importância da informação no contexto da sociedade atual, uma sociedade que vive em rede, principalmente no que diz respeito ao conhecimento e que é denominada por muitos como a Sociedade da Informação. Ou seja, o conceito de rede proposto principalmente pela Web é uma realidade exercendo influências em todos os outros setores da sociedade, seja na economia, na saúde, na segurança, na educação, entre outros. Segundo Morin (2003a, p.109), “não aplicamos os esquemas tecnológicos apenas ao trabalho manual ou mesmo à máquina artificial, mas também às nossas próprias concepções de sociedade, vida e homem”, configurando a materialização do que Pierre Lévy chama de “espaço do saber” (LÉVY, 1999, p.24).

Nesse sentido de um “saber coletivo”, Freire (2005) resgata a visão de Paul Otlet e Henri de La Fontaine, evidenciada por Mattelart (2006) e Lévy (1999), desvinculando a informação das instituições, mas colocando-a como algo em benefício da sociedade. Tem-se, então, a “origem da CI, no contexto da emergência do próprio espaço do saber” (FREIRE, 2005, p.137). Estabelece-se, pois, através dessa última afirmação, o primeiro paralelo entre a CI e a Web, pois aqui se entende a primeira como o resultado da evolução da proposta de Otlet e La Fontaine; e a segunda como o possível e, aparentemente, principal meio de materialização desse alcance global do conhecimento.

O presente trabalho aborda a Web, em sua proposta atual, denominada de Web Semântica (WS), à luz da CI, pois, apesar das indicações nos parágrafos anteriores, percebe-se que efetivamente a Web atual funciona em um “caos informacional”, devido à descentralização total de inserção de conteúdos, com critérios geralmente individuais na estruturação da representação dos conteúdos, acompanhada de limitações nas TICs, haja vista, ser a Web uma tecnologia que se popularizou por volta de 1990, ou seja, uma tecnologia nova, que está em pleno desenvolvimento.

Esses problemas podem ser exemplificados de dois modos na Web atual. Em primeiro lugar, através dos mecanismos de busca, em que a atividade de busca se realiza apenas por palavras-chave, sem

⁴ Dimensão ou domínio virtual da realidade, constituído por entidades e ações puramente informacionais; meio, conceitualmente análogo a um espaço físico, em que seres humanos, máquinas e programas computacionais interagem (FERREIRA, 2004); novo meio de comunicação que surge da interconexão mundial de computadores; não apenas a infra-estrutura material da comunicação digital, mas também o universo de informações que ela abriga. (LÉVY, 2003, p.17).

⁵ Conjunto de técnicas (materiais e intelectuais), de práticas, de atitudes, de modos de pensamento e de valores que se desenvolvem juntamente com o crescimento do ciberespaço. (LÉVY, 2003, p.17).

organização semântica alguma, ou seja, a informação levantada normalmente possui pouca relevância para o usuário. Em segundo lugar, a forma de construção dos *sites*⁶ na Web atual utiliza uma linguagem de marcação denominada HTML⁷, que tem como objetivo principal apenas a apresentação (interface) do conteúdo para o usuário final, não havendo preocupação com a padronização e a estruturação dos dados para futura recuperação dessa informação.

Diante desse cenário, a WS surge com uma proposta de melhorar a organização da informação no ambiente Web, ou seja, não se trata de uma nova Web, mas uma evolução da atual, que não nega e nem contraria a descentralização na inserção de conteúdos, mas apresenta mecanismos que possibilitam diminuir a passividade entre usuários e máquina, aumentando a interação desses, ou seja, basicamente a WS trata da “criação e implantação de padrões (standards) tecnológicos para permitir este panorama” (SOUZA; ALVARENGA, 2004, p.134), além de um tratamento da informação com uma significação clara, para que haja efetivamente a interação entre os atores citados (BERNERS-LEE; HENDLER; LASSILA, 2001).

Os problemas da Web não são objeto de apenas uma área do conhecimento, mas no que diz respeito à organização da informação, RI, inserção de conteúdos, estruturação de representação de conteúdos, apenas mencionando alguns aspectos, trata-se de um ambiente de estudo e de implantação de métodos da CI.

Assim, esse trabalho se justifica em dois momentos. Primeiro, entende-se que a WS une dois pontos considerados por Smit (1986) como fundamentais na troca do paradigma do acervo para o paradigma da informação, itens que corroboram o surgimento e desenvolvimento da CI: linguagem e novas tecnologias.

Como o tratamento da informação é de competência (não exclusiva) da CI, se torna importante essa aproximação com a WS. O conhecimento da CI pode ajudar na construção e manutenção dessa tecnologia, principalmente no que diz respeito à construção de “representações da informação” como os sistemas de organização e classificação, o estabelecimento de metadados, além do estudo de usuário, com uma abordagem que o valoriza como principal ator do ciclo da informação, pois pela sua ação ocorre a RI.

A teoria do conhecimento objetivo de Popper (1975) é o segundo item que justifica o trabalho. Essa teoria descreve três mundos, sendo o primeiro o mundo físico (matéria, as coisas), o segundo mundo

⁶ Conjunto de documentos apresentados ou disponibilizados na Web por um indivíduo, instituição ou empresa, e que pode ser fisicamente acessado por um computador e em endereço específico da rede. (FERREIRA, 2004).

⁷ “Hypertext Markup Language” – Linguagem de marcação hipertextual – uma coleção de comandos de formatação que criam documentos hipertextuais ou, mais simplesmente, páginas da Web. (COSTA, 2003).

é o subjetivo (pensamentos, o entendimento (classificação) do homem em relações aos objetos do primeiro mundo) e o terceiro mundo, denominado objetivo, que é o mundo do conhecimento registrado, que permite a inteligibilidade dos três mundos e toda a atividade científica. Sobre o terceiro mundo ou “*mundo 3*” ou mundo do conhecimento objetivo, Popper (1975, p.78) diz que exemplos desse *mundo* “são as teorias publicadas em revistas e livros e conservadas em bibliotecas; discussões dessas teorias; dificuldades ou problemas apontados em conexão com essas teorias; [...] mundo dos conteúdos lógicos de livros, bibliotecas, memórias de computador e similares [...]”.

Podemos afirmar que esse conhecimento que está armazenado em bibliotecas ou em outras unidades de informação, está também sendo armazenado na Web, de forma crescente e até prioritária, pois, além de iniciativas exclusivas para a Web, como revistas científicas eletrônicas, há também um crescente movimento de atividades que eram exclusivamente não digitais, como a publicação de revistas científicas impressas, que agora possuem uma versão eletrônica. Esse exemplo pode ser estendido ao comércio, ao governo, etc.

São três as teses de Popper (1975, p.78) sobre o “*mundo 3*”:

- (1) Podemos descobrir no mundo 3 problemas novos que lá estavam antes de ser descobertos e antes mesmo de se tornarem conscientes [...];
- (2) Assim, há um sentido em que o mundo 3 é autônomo: neste mundo podemos fazer descobertas teóricas de modo semelhante àquele por que fazemos descobertas geográficas no mundo 1;
- (3) quase todo o nosso conhecimento subjetivo (conhecimento do mundo 2) depende do mundo 3, isto é, de teorias formuladas lingüisticamente [...].

Numa aproximação da CI com essa teoria, Miranda (2003) diz que o “*mundo 3*” é o campo de atuação da área, sendo fundamental para o desenvolvimento da sociedade. Assim, parece necessário um estudo para entender essa nova proposta da Web e indicar alguns pontos congruentes com a CI. Portanto, definem-se a seguir os objetivos da presente pesquisa.

1.1 Objetivos

Objetivo geral

- Indicar perspectivas na WS para a CI, principalmente no que diz respeito à organização e à RI.

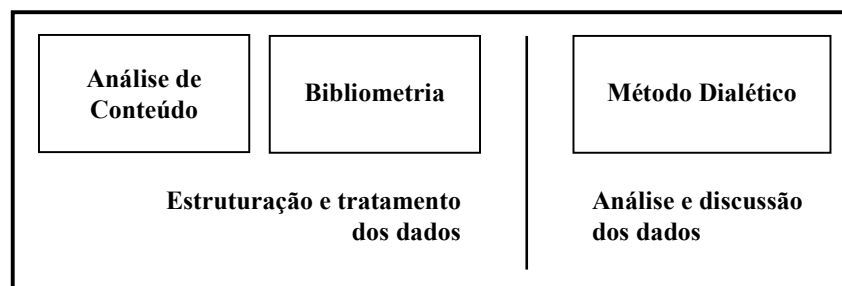
Objetivos específicos

- Identificar pontos de desenvolvimento da WS, principalmente conceituais, a partir de uma literatura representativa e pertinente sobre o assunto, escolhidas conforme critérios estabelecidos pela análise de conteúdo (AC).
- Indicar como a WS se integra à literatura da CI no Brasil, através das teses e das dissertações dos programas de Pós-Graduação em CI.

1.2 Procedimentos metodológicos

O núcleo da pesquisa teve como objeto de estudo os aspectos conceituais que relacionam a CI e a WS, na busca de responder aos objetivos propostos. O método utilizado foi o dialético, com abordagem exploratória e descritiva. O procedimento técnico utilizado para a realização da pesquisa foi o bibliográfico e as técnicas para a estruturação e apresentação dos materiais bibliográficos levantados, foram a bibliometria e a AC. A Figura 1 abaixo ilustra o método dialético em relação aos procedimentos técnicos de estruturação e tratamento dos dados.

Figura 1. Relação do método dialético com as técnicas de estruturação e tratamento dos dados.



1.2.1 Método dialético

A abordagem dialética parece ser a que melhor se encaixa como método de estudo do tema proposto, pois busca a interpretação da realidade de forma completa. Demo (1983, p.85) diz que é “a metodologia mais correta para as ciências sociais, porque é aquela que, sem deixar de ser lógica, demonstra sensibilidade pela face social dos problemas”, porém, adverte: “é preciso, contudo, reconhecer, por coerência dialética, que as estruturas identificadas na realidade são hipóteses de trabalho, não afirmações perenes” (DEMO, 1983, p.92).

Essa proposta metodológica surge na antiguidade, sendo basicamente a arte de dialogar, mas logo passa a ser “a capacidade de demonstrar uma tese através de uma argumentação, definindo e distinguindo com clareza os conceitos envolvidos” (KONDER, 1998, p.7), ou seja, um modo específico de argumentar, buscando as contradições – análise e síntese, indica Gadotti (1984), afirmando ainda, assim como Demo (1983), ser a contradição uma das principais características desse método.

Konder (1998, p.49) diz que o conceito de contradição adotado pela dialética não pode ser confundido com um “mero defeito”, mas “[...] num sentido amplo, filosófico, que não se confunde com o sentido que a lógica confere ao termo, a contradição é reconhecida pela dialética como princípio básico de movimento pelo qual os seres existem”, sendo sua marca registrada, pois a contradição é uma característica da realidade e não um defeito (DEMO, 1983). Assim podemos afirmar que essa contradição se relaciona à mudança, alteração, relação, pois “a dialética considera todas as coisas em movimento, relacionadas umas com as outras” (GADOTTI, 1984, p.16). A dialética, portanto, privilegia o fenômeno da transição histórica, o que significa uma superação de uma fase por outra, predominando mais elementos do novo do que repetições possíveis da fase anterior (DEMO, 1983), ou seja, uma indicação clara de evolução.

Mas para que haja contradição, é necessário sempre um referencial, por isso a idéia de dialogar. “Heráclito de Éfeso indica ser a dialética a luta dos opostos: frio-calor, vida-morte [...]” (GADOTTI, 1984, p.15), ou seja, definimos a morte a partir da existência da vida, sendo assim, um não existe sem o outro. Assim, “segundo Aristóteles, todas as coisas possuem determinadas potencialidades; os movimentos das coisas são potencialidades que estão se atualizando, isto é, são possibilidades que estão se transformando em realidades efetivas” (KONDER, 1998, p.10).

Para um melhor entendimento desse método recorre-se a três termos-chave que a caracterizam: tese,

antítese e síntese. A tese significa quaisquer formações sociais, vigentes na história, a antítese é naturalmente a contradição da tese, e a síntese é a superação das duas idéias anteriores em uma nova tese, que possui naturalmente sua contradição, e isso forma um ciclo contínuo (DEMO, 1983).

Mandel (1978, p.116)⁸ citado por Gadotti (1984, p.27), diz que a dialética pode ser dividida em três níveis: da natureza, da história e do conhecimento. Este estudo se baseará na dialética do conhecimento que “é uma dialética sujeito-objeto”, onde há “o resultado de uma interação constante entre os objetos a conhecer e a ação dos sujeitos que procuram compreendê-los”. Sendo assim “através do método dialético o fenômeno ou coisa estudada deverá apresentar-se ao leitor de tal forma que ele o apreenda em sua totalidade” (GADOTTI, 1983, p.31). Para tanto, são três as leis da dialética que Konder (1998, p.58) apresenta:

- 1) passagem da quantidade à qualidade (as coisas não mudam sempre no mesmo ritmo, o processo de transformação por meio do qual elas existem passa por períodos lentos – alterações quantitativas – e por períodos de aceleração – alterações qualitativas, modificações radicais);
- 2) interpenetração dos contrários (tudo tem a ver com tudo, os diversos aspectos da realidade se entrelaçam e, em diferentes níveis, dependem uns dos outros); e
- 3) lei de negação da negação, ou seja, o movimento da realidade não se esgota em contradições sem sentido, nem na repetição de conflitos entre teses e antíteses, entre afirmações e negações. A afirmação e a negação são superadas, e prevalece uma nova síntese; é a negação da negação.

Engels (1976)⁹ citado por Gadotti (1984, p.24) explica da seguinte forma as leis citadas:

A primeira significa que na natureza as variações qualitativas podem ser obtidas somente acrescentando-se ou tirando-se matéria ou movimento por meio de variações quantitativas. A segunda garante a unidade e a continuidade da mudança incessante na natureza e nos fenômenos. A terceira garante que cada síntese é por sua vez a tese de uma nova antítese reproduzindo indefinidamente o processo.

Para Gadotti (1984, p.24) há quatro leis da dialética, que ele prefere chamar de princípios. Partindo das leis já citadas e acrescentando uma, o autor explica de forma mais sucinta tais postulados da dialética. São elas:

- 1) tudo se relaciona (princípio da totalidade): a natureza se apresenta como um todo coerente.

⁸ MANDEL, E. **Introdução ao marxismo**. Porto Alegre: Movimento, 1978.

⁹ ENGELS, F. **Dialética da natureza**. Rio de Janeiro: Paz e Terra, 1976.

Levando em conta essa ação recíproca, examina os objetos e fenômenos buscando entendê-los numa totalidade concreta – Lei de interpenetração dos opostos;

2) tudo se transforma (princípio do movimento): como o movimento é inerente a todas as coisas, a afirmação e a negação são superadas pela síntese, ou seja, uma nova tese, é a negação da negação;

3) mudança qualitativa (princípio da mudança qualitativa); e

4) unidade e lutas dos contrários (princípio da contradição) – princípio fundamental da dialética. “A existência dos contrários não é um absurdo lógico, ela se funda no real”.

Segundo Demo (1983, p.98), as características dessa metodologia são muitas. Algumas são as seguintes: a) problematiza com maior inteligência a relação entre sujeito e objeto, superando posições estanques e estereotipadas ligadas a visões estáticas da objetividade e da neutralidade; b) através da concepção da identidade de contrários, adapta-se melhor à dinâmica histórica; c) privilegia os fenômenos da transição histórica, ainda que possa ver os outros também; d) propõe a visão de totalidade, no sentido de esforçar-se por recortar menos a realidade e de não formalizá-la em partes estáticas, pois a visão das partes segmentadas não é dinâmica; e) é a metodologia mais crítica e autocrítica, como requer a dinâmica da realidade social.

Lefèbvre (1975, p.241)¹⁰ citado por Gadotti (1984, p.33) indica algumas “regras práticas” para o método dialético, a saber: 1) analisar de forma objetiva o objeto; 2) apreender o conjunto das conexões internas, dos aspectos, do desenvolvimento e do movimento do objeto; 3) apreender a coisa na sua totalidade e unidade dos contrários; 4) analisar a tendência (o que tende a ser e o que tende a cair no nada); 5) “não esquecer de que tudo está ligado a tudo; e que uma interação insignificante, negligenciável por não ser essencial em determinado momento, pode tornar-se essencial num outro momento ou sob um outro aspecto”; 6) captar as transições; 7) jamais estar satisfeito com o obtido; 8) apreender conexões e movimentos; 9) superar, modificar, mas por vezes voltar, rever, repetir, e até mesmo retornar ao início.

Assim, a partir do que foi apresentado pretende-se, procurou-se demonstrar os seguintes movimentos e relações para o alcance dos objetivos propostos:

1. Web atual *versus* WS (movimento, mudança qualitativa);
2. O movimento inicial até o estado atual da WS (mudança qualitativa);
3. CI (teses e dissertações) *versus* WS (relações); e

¹⁰ LEFÈBVRE, H. **Lógica formal, lógica dialética**. Rio de Janeiro: Civilização Brasileira, 1975.

4. CI (revisão de literatura) *versus* WS (relações).

1.2.2 Abordagem metodológica exploratória e descritiva

Para entender aspectos atuais da WS, foram identificadas aplicações para métodos da CI na WS e como a CI tem estudado a WS. O tipo de pesquisa que foi desenvolvido se caracteriza por ser exploratório, pois existe a necessidade do “aprimoramento de idéias ou a descoberta de intuições” (GIL, 2002, p.41), que “aumenta a familiaridade do pesquisador com o fenômeno ou com o ambiente que pretende investigar, servindo de base para uma pesquisa futura mais precisa” (DENCKER; VIÁ, 2002, p.59). Reforça ainda essa abordagem metodológica exploratória, a seguinte afirmação sobre tal procedimento, que:

[...] pode utilizar métodos quantitativos e qualitativos: revisão da literatura, pesquisa bibliográfica e documental para elaboração de resenha da ciência social afim, assim como de outras partes pertinentes da literatura que tenham relação com o objeto que se pretende estudar. [...] análise de exemplos que estimulem a compreensão. (DENCKER; VIÁ, 2002, p.59).

O tipo de pesquisa descritiva também se aplica na busca dos objetivos propostos, pois se trata de um tipo que visa a descrição das características de determinado fenômeno, além de procurar descobrir a possível existência de associações entre certas variáveis. (GIL, 2002).

1.2.3 Pesquisa bibliográfica

O procedimento técnico adotado para realização da pesquisa foi o bibliográfico, que se constitui do levantamento de materiais como livros e artigos científicos (GIL, 2002) e também teses e dissertações, pois a partir desse conjunto de material que Popper (1975) chama de autônomo, pode-se fazer descobertas sobre um determinado objeto ou mesmo procurar entendê-lo. Assim, foram delimitados os termos para a pesquisa bibliográfica, conforme indica a Quadro 1, a partir de consulta

aos tesauros das bases de dados LISA (Library and Information Science Abstracts) e Web Wilson, com acesso pelo portal da Pontifícia Universidade Católica de Campinas.

Quadro 1. Termos-chave para pesquisa bibliográfica

Termos primários	Termos secundários
<ul style="list-style-type: none"> • Web Semântica • Ontologia • World Wide Web • Metadados • Recuperação da Informação na Web/Internet 	<ul style="list-style-type: none"> • Taxonomia • Tesauro • Vocabulário controlado • XML – Linguagem de marcação • Catalogação na Internet • Indexação automática

O termo principal é Web Semântica. Para os outros termos existe a necessidade de uma combinação direta ou mesmo indireta com o termo principal.

1.2.4 Técnicas metodológicas para estruturação e apresentação dos dados

Para efeito de análise do material bibliográfico levantado, foram utilizadas duas técnicas metodológicas, apresentadas a seguir: a bibliometria e a AC.

1.2.4.1 Bibliometria

O uso da matemática e da estatística na CI remonta aos anos 20 do século XX através da iniciativa de Alfred Lotka em 1926. Outros estudiosos o seguiram. Foram eles: Samuel C. Bradford em 1934, George Zipf em 1935, Claude Shannon em 1948, Philip M. Morse em 1968 e Derek J. de Solla Price em 1976. Todos eles são mencionados por Le Coadic (2007, p.221) com exemplos de usos das disciplinas citadas que influenciaram a área da informação, pois “confirmam a existência, [...] de regularidades, distribuições e relações mensuráveis universais [...]”.

A aplicação da matemática na CI se manifesta em três níveis segundo Le Coadic (2007), sendo eles a: 1) capacidade preditiva: sugere a realização de observações ou experimentações e fornece resultados numéricos; 2) capacidade retroditiva: reproduz resultados já conhecidos e os organiza em um formalismo conciso; e 3) capacidade explicativa: fornece uma explicação de fenômenos, através de uma sequência de inferências relacionando suas descrições a princípios reconhecidos como fundamentais. Assim, a matemática se apresenta na CI como “uma linguagem capaz de descrever, explicar e dominar os fenômenos” (LE COADIC, 2007, p.226).

A estatística é um ramo da matemática que lida com a análise dos valores numéricos, que pode ser descritiva ou interpretativa (LE COADIC, 2007, p.230). “No plural, as estatísticas significam uma coleção de valores numéricos relativos a uma categoria de fatos e objetos [...]” (LE COADIC, 2007, p.232).

Na área da CI, para o uso da matemática e estatística cunha-se o termo bibliometria, que se ocupa da medida ou quantificação aplicada aos livros ou documentos. Trata-se do método científico de uma disciplina científica proposta por Paul Otlet, chamada Bibliologia, “uma ciência geral que compreende o conjunto sistemático dos dados relativos à produção, conservação, circulação e uso dos escritos e dos documentos de toda espécie” (OTLET, 1986, p.20).

Pritchard (1969)¹¹ citado por Urbizagástegui Alvarado (2007, p.197) é oficialmente o primeiro a utilizar o termo, definindo a bibliometria como a “aplicação de modelos matemáticos e estatísticos aos livros e a outros meios de comunicação escrita”.

Brookes (1973)¹² citado por Urbizagástegui Alvarado (2007, p.198) estabeleceu, para as técnicas bibliométricas, as seguintes utilidades:

1. no desenho de sistemas de informação mais econômicos; 2. na melhoria da eficiência dos processos de gerenciamento de informação; 3. na identificação e medidas nos atuais serviços bibliográficos; 4. na predição das tendências de publicação; e 5. no descobrimento e elucidação de leis empíricas que poderiam formar a base de desenvolvimento de uma teoria da Ciência da Informação.

Para a aplicação das técnicas bibliométricas é necessário o que Le Coadic (2007) chama de conjuntos, ou seja, um grupo de material bibliográfico onde se aplicam técnicas de medição. Para o

¹¹ PRITCHARD, A. Statistical bibliography or bibliometrics? *Journal of Documentation*, London, v.25, n.4, p.348-349, Dec. 1969.

¹² BROOKES, B. C. Numerical methods of bibliographical analysis. *Library Trends*, Champaign, v.22, n.1, p.18-43, July 1973.

presente trabalho haverá dois tipos de conjuntos: 1) qualquer tipo de material bibliográfico sobre a WS; e 2) teses e dissertações de programas de pós-graduação em CI do Brasil.

Foi aplicada a bibliometria ao conjunto de teses e dissertações com o objetivo de quantificar as palavras-chave e suas relações em cada trabalho analisado (exemplo: com quais outras palavras-chave o termo ‘*web semântica*’ aparece), assim como a autoria, os orientadores, as linhas de pesquisa, a formação em nível de graduação de cada autor (informação que será retirada da Plataforma Lattes, nos respectivos currículos), etc. A análise foi unidimensional, ou seja, sem fazer relações profundas de inferências a partir dessa técnica, pois não houve necessidade para o alcance dos objetivos. Para os demais materiais bibliográficos utilizados para a apresentação do panorama da WS, não foram aplicadas essas diretrizes bibliométricas. Foram usados os critérios de representatividade e de pertinência retirados da AC, para entender esse fenômeno.

1.2.4.2 Análise de conteúdo

A AC, segundo Bardin (2000, p.31), consiste em um “conjunto de técnicas de análises de comunicações”. Assim “tudo o que é dito ou escrito é susceptível de ser submetido a uma AC” (HENRY; MOSCOVIEI, 1968¹³ citados por BARDIN, 2000, p.33), que objetiva ultrapassar a incerteza do conteúdo de alguma mensagem e enriquecer a leitura.

Um domínio possível para a aplicação das técnicas de AC é o suporte escrito do código lingüístico, mais precisamente em comunicação de massa como jornais, livros, anúncios publicitários, cartazes, literatura, textos jurídicos, panfletos, etc. (BARDIN, 2000). Para um melhor entendimento da AC, no Quadro 2 são apresentados outros modos de lidar com a linguagem e os documentos.

¹³ HENRY, P.; MOSCOVIEI, S. Problèmes de l’analyse de contenu. **Langages**, Paris, n.2, set.1968.

Quadro 2. Especificidades da análise de conteúdo

	Especificidade	Especificidade da AC
Linguística	Tem por objetivo o aspecto coletivo e virtual da linguagem; sua teoria.	Trabalha a palavra, a prática da língua realizada por emissores identificáveis.
Semântica	Estuda os sentidos das unidades linguísticas, sempre de forma universal.	Busca os sentidos das palavras também, inclusive podendo utilizar a semântica, mas sempre de forma contextual e dirigida.
Análise documental	Busca representar o conteúdo de um documento, para que o mesmo seja posteriormente recuperado.	Não visa o documento e sim a manipulação das mensagens contidas em um ou mais documentos, com o objetivo de evidenciar os indicadores presentes nos documentos.
Análise do discurso (AD)	Analisa em que perspectiva a relação social de poder se constrói no plano discursivo, vai além do texto, considerando o contexto sócio-histórico	Busca o saber a partir da superfície textual, se limitando ao que o texto quer dizer.

Fonte: Adaptação feita a partir de Bardin (2000), Mendonça (2007) e Rocha; Deusdará (2005).

Percebe-se, portanto, que a AC tem sua especificidade em relação às outras formas mencionadas de trabalhar a linguagem. Enquanto a Linguística estuda a linguagem não balizada especificamente por um documento, se manifestando *‘de fora para dentro’*, ou seja, do contexto social para os documentos, a AC trabalha a linguagem *‘de dentro para fora’*, isto é, parte dos documentos para entender a mensagem. Quanto à semântica, ela se refere a uma subárea da linguística, estudando os sentidos dos enunciados de forma geral, ou seja, também *‘de fora para dentro’*, tendo base o social e os documentos, já a AC parte dos documentos para identificar os sentidos das mensagens, sempre com um objetivo específico. Parecem claras, pelo quadro, as diferenças entre a AC e a análise documental, cuja preocupação é a representação do conteúdo de um documento para recuperação, diferente da AC que procura demonstrar a mensagem do documento. Em relação à análise do discurso (AD), as diferenças já não são tão evidentes, Bardin (2000) afirma que essa é uma técnica da AC, mas a diferença está também condicionada ao documento e ao sujeito que emite a mensagem. A AC é restrita ao documento, a AD é contextual, buscando a posição de um sujeito sobre um determinado assunto, sendo mais subjetiva; é um discurso social, por isso usada para a análise de comunicação oral de um público específico.

Segundo Caregnato e Mutti (2006), enquanto a AC pode ter caráter quantitativo e qualitativo, a AD é essencialmente qualitativa, pois não se prende ao texto, mas ao sentido da mensagem de algum emissor, ou seja, a condição do sujeito que emite tal mensagem é essencial para a AD, por isso é um

método muito usado para trabalhos de saúde pública. Já a AC para Orlandi (1987)¹⁴ citado por Mendonça (2007, p.155), apresenta as seguintes características: 1) o texto é um documento restrito, limitado ao seu próprio contexto; 2) parte do discurso para a enunciação; 3) visa o que o texto quer dizer; 4) busca o sentido do texto expresso em sua estrutura e o interpreta a partir daí; e 5) não há relação direta com a historicidade. Ela não vai além do texto, ou seja, não realiza relações além deste, mas mesmo assim pode ter caráter qualitativo, desde que considere a identificação das características da mensagem, sem levar em conta a quantificação (CAREGNATO; MUTTI, 2006).

A AC apresenta as seguintes etapas segundo Bardin (2000): organização da análise, codificação, categorização, inferência e tratamento informático.

A organização da análise está dividida em três fases. A primeira é a pré-análise. Ela consiste na escolha dos documentos a serem analisados e parte de uma leitura que Bardin (2000) denomina flutuante, é o primeiro contato com os documentos. Em seguida há a escolha propriamente dita dos documentos, que pode ocorrer em função de um objetivo estabelecido ou de um universo documental determinado sem um objetivo definido. As principais regras para a escolha dos documentos, segundo Bardin (2000), são:

- Regra de exaustividade: dependendo da necessidade em questão, nenhum documento pode ficar de fora da análise;
- Regra de representatividade: desde que a necessidade de análise permita pode-se utilizar uma amostra de material;
- Regra de homogeneidade: os documentos devem apresentar critérios de homogeneidade, a partir dos objetivos ou hipóteses estabelecidas; e
- Regra de pertinência: o material deve ser adequado, enquanto fonte de informação.

Para o presente trabalho, a partir dos conjuntos bibliográficos já citados, as regras de representatividade e de pertinência se aplicam para a seleção da literatura em geral sobre a WS (1º objetivo específico) e regras de exaustividade e de homogeneidade se aplicam a seleção das teses e dissertações (2º objetivo específico), vale salientar porém, que para esse segundo conjunto de documentos não foi aplicada a AC.

O terceiro item da pré-análise é a formulação das hipóteses ou dos objetivos. A primeira é uma questão que baliza as análises, e a segunda é a finalidade geral da análise. Vale salientar que há

¹⁴ ORLANDI, E. **A linguagem e seu funcionamento**: as forças do discurso. Campinas: Pontes, 1987.

análises sem hipóteses. O quarto item da pré-análise é a referenciação dos índices e a elaboração de indicadores. Um exemplo, pode ser o número de vezes que um termo ou palavra está contido em um documento, se ele é freqüente pode ter muita relevância, caso não apareça, pode haver pouca relevância. A preparação ou edição do material é a última parte da pré-análise, consiste em deixá-lo em uma forma em que deverá ser mais facilmente entendido. Pode ser exemplificado pelo recorte de artigos de imprensa ou pela anotação em fichas de respostas a questões abertas, pois dessa forma tal material estará em uma forma melhor para a análise (BARDIN, 2000).

No presente trabalho não há hipóteses, sendo assim todo o material foi estudado num sentido de estruturar e apresentar todas as informações possíveis, a partir dos objetivos estabelecidos.

As duas últimas etapas da organização da análise são: a exploração do material (que é a sequência das decisões estabelecidas na pré-análise, em função das regras já formuladas); e o tratamento dos resultados obtidos e interpretação (consiste em dar significado aos dados brutos alcançados, ou seja, desenvolver quadros de resultados, figuras, modelos, diagramas, que podem simplificar e apresentar as informações fornecidas pela análise) (BARDIN, 2000).

Após a organização, o próximo passo é realizar a codificação do material, que se trata do “processo pelo qual os dados brutos são transformados sistematicamente e agregados em unidades, as quais permitem uma descrição exata das características pertinentes do conteúdo” (HOLSTI, 1969¹⁵ citado por BARDIN, 2000, p.103). Para uma análise quantitativa e categorial, a organização da codificação consiste em três passos, segundo Bardin (2000): o recorte (escolha das unidades); a enumeração (escolha das regras de contagem); e a classificação e agregação (escolha das categorias).

O recorte pode se basear em duas unidades: a unidade de registro e a unidade de contexto. A unidade de registro é a unidade base de significação que se deve codificar, o recorte nessa unidade pode ser em nível semântico (o tema) ou lingüístico (palavra ou frase). As unidades de registro mais utilizadas, segundo Bardin (2000), são: a palavra; o tema (uma afirmação acerca de algum assunto, um conceito; é complexa e varia em comprimento); o objeto ou referente (temas-eixo que conduzem o discurso); o personagem (obras de ficção podem ser analisadas em função de um personagem); o acontecimento (relatos ou narrações conduzem a análise); e o documento (desde que possa ser caracterizado globalmente e no caso de análise rápida).

A unidade de recorte de contexto serve de compreensão para a codificação da unidade de registro, pois uma palavra, um tema, um objeto, um personagem, um acontecimento e um documento são

¹⁵ HOLSTI, O. R. **Content analysis for the social sciences and humanities**. [S.l.]: Addison-Wesley, 1969.

unidades que variam em relação ao contexto (BARDIN, 2000).

A unidade de recorte de registro utilizada para o presente trabalho foi o tema, pois tal unidade pode identificar as propostas de cada material estudado. Quanto à unidade de recorte de contexto, ela está balizada na CI e nos processos de representação e de RI.

As regras de enumeração consistem no modo de contagem da unidade de registro, e segundo Bardin (2000) se dividem em: a presença ou a ausência (unidades de registro que estão presentes em uma análise têm algum significado, assim como sua ausência também); a frequência (a importância de uma unidade de registro aumenta conforme sua frequência); a frequência ponderada (consiste em determinar pesos às unidades de medida); a intensidade (que pode ser exemplificada pelo tempo de uso de um verbo ou pela designação de atributos qualificáveis); a direção (ponderação de frequência qualitativa, ou seja, algo mais subjetivo no momento de medir, como bonito/feio – critério estético, ou pequeno/grande – tamanho); a ordem (consiste na ordem de aparecimento de uma unidade de registro); e a co-ocorrência (a presença de duas ou mais unidades de registro em uma unidade de contexto).

Na AC, a enumeração faz uso da matemática e da estatística. Há certa semelhança com a bibliometria. No entanto, a primeira está focada no conteúdo do texto. Para efeito desse trabalho, a bibliometria vai ser aplicada na análise de diversas informações sobre as teses e dissertações dos programas de pós-graduação em CI no Brasil, geralmente voltadas para a identificação dos materiais e seus objetivos, e a AC vai avaliar aspectos do texto em si.

A proposta de enumeração é mais adequada para as unidades de recorte de registro para palavra, tema e tema-eixo (ou objeto), e responde a questões mais quantitativas, sendo assim, essa etapa não foi utilizada nesse trabalho.

A categorização é a terceira etapa da AC proposta por Bardin (2000) e consiste na operação de classificar os elementos que constituem o conjunto de documentos, diferenciando-os e, em seguida, os reagrupando. Nesse processo há duas etapas: o inventário (isolamento dos elementos); e a classificação (divisão dos elementos para que haja organização das mensagens).

A qualidade no processo de categorização é fundamental, sendo assim, um conjunto de categorias deve possuir as seguintes qualidades: a exclusão mútua (um elemento não pode estar em mais de uma divisão, ou seja, não pode ser classificado em mais de duas categorias); a homogeneidade (um único princípio de classificação deve balizar a organização, para que haja a exclusão mútua); a pertinência (o sistema de categorias deve refletir as intenções da análise); a objetividade e a

fidelidade (diferentes partes de um certo material aplicadas às mesmas categorias devem ser codificadas da mesma maneira); e a produtividade (as categorias devem produzir resultados) (BARDIN, 2000). A categorização foi realizada a partir dos resultados de análises do material base sobre a WS, textos pioneiros que indicam quais os objetivos, características e fundamentos dessa nova Web. O Quadro 3 indica quais foram as categorias utilizadas nas análises da pesquisa.

Quadro 3. Categorias para a análise de conteúdo

CATEGORIAS	DESCRIÇÃO
Conceitos/características	O conceito é a representação de um objeto em estudo pelo pensamento, sendo assim, ele o define e o caracteriza. As características que são particularidades de um objeto acabam por defini-lo.
Objetivos	Proposta alvo, foco de atuação de um objeto.
Estrutura	Partes que compõem o todo.
Modelos conceituais utilizados	Representação conceitual que pode em certa medida, influenciar ou balizar algo real/existente, como o hipertexto, que foi utilizado na criação da Web.
Ferramentas tecnológicas	Recurso empregado para se alcançar um objetivo, conseguir um resultado.
Cenários	Exemplos de aplicação do objeto.

A inferência é a próxima etapa do método de AC e consiste basicamente em tirar conclusões de todo material organizado, devendo haver pólos que direcionem tais inferências (BARDIN, 2000). O processo de inferência do presente trabalho foi realizado a partir da criação das categorias, do tema como unidade de recorte de registro, assim como da unidade de recorte de contexto, a CI, através da representação do conhecimento e da informação, além da RI.

O tratamento informático é a última etapa. Consiste na utilização de algum software para auxiliar na estruturação e inferência dos dados. Geralmente é necessário para a criação das representações gráficas quando há grande volume de dados. Para essa pesquisa essa etapa não foi necessária.

O Quadro 4 indica quais dessas etapas descritas acima foram utilizados no presente trabalho.

Quadro 4. Etapas da análise de conteúdo utilizadas no trabalho

ETAPAS DA ANÁLISE DE CONTEÚDO	Utilizadas
1. Organização da análise	X
A. Pré Análise (Leitura flutuante)	X
Leitura flutuante	X
Escolha dos documentos	X
Formulação das hipóteses ou objetivos	Não
Referenciação dos índices e elaboração de indicadores	Não
Preparação ou edição do material	X
B. Exploração do material (aplicação do que foi estabelecido na pré-análise)	X
C. Tratamento dos resultados e interpretação	X
2. Codificação	X
A. Recorte (escolha das unidades)	X
Unidade de registro	Não
Palavra	Não
Tema (um conceito)	X
Objeto ou referente (tema-eixo)	Não
Personagem	Não
Acontecimento	Não
Documento	Não
Unidade de contexto	X
B. Enumeração (regras de contagem)	Não
Presença ou ausência da unidade	Não
Frequência e frequência ponderada	Não
Intensidade e direção	Não
Ordem e co-ocorrência	Não
3. Categorização	X
A. Inventário (isolar elementos)	X
B. Classificação (dividir elementos para ordenar a mensagem)	X
4. Inferência	X
5. Tratamento informático	Não

O processo de busca dos temas aconteceu a partir do isolamento dos parágrafos de cada texto analisado, conforme indica o exemplo no Quadro 5 abaixo. Buscou-se em cada parágrafo frases

significativas para os objetivos propostos. Entende-se parágrafo como unidade de texto escrito cuja função é a de indicar o desenvolvimento de uma idéia, de um conceito, de um tema, através das frases que o formam.

Quadro 5. Exemplo da extração dos temas de um parágrafo

A Web continua a crescer em ritmo assustador, estima-se que já ultrapassamos oito bilhões de páginas.
No entanto, grande parte das páginas disponíveis na Web ainda mantém muito de sua característica
inicial, ou seja, são direcionadas para outras pessoas e não para serem processadas por programas
de software . Computadores são usados meramente para mostrar a informação na tela , ou seja,
decodificar as marcações de cores, posição e links, codificadas através das linguagens HTML ou
XML .

Fonte: Breitman (c2006, p.2).

A partir do destaque no quadro acima, temos quatro temas (conceitos, idéias) isolados: 1) páginas web direcionadas para pessoas; 2) não para serem processadas por programas de software; 3) computadores usados apenas para mostrar a informação na tela; e 4) codificação ocorre através das linguagens de marcação HTML e XML¹⁶. A título de exemplo, podemos considerar que os três primeiros temas caracterizam e conceituam a web atual, sendo assim, eles foram colocados na categoria “conceitos/características”. O quarto tema isolado indica duas ferramentas utilizadas para a realização das atividades de construção da web e foi colocado na categoria “ferramentas tecnológicas”.

1.4 Apresentação dos capítulos

O capítulo introdutório apresenta a contextualização, os objetivos, a justificativa e os procedimentos metodológicos utilizados do trabalho. Os demais capítulos são:

- Capítulo 2: descreve a CI, apresentando seu contexto científico, origem e períodos de desenvolvimento, caracterizando a biblioteconomia, a documentação e as TICs, como relações

¹⁶ Sigla de *eXtensible Markup Language*, linguagem de marcação (ASSOCIAÇÃO..., 2007).

que colaboram para a origem da área. Em seguida apresentam-se alguns conceitos de CI e o seu objeto de estudo, a informação. Encerra-se o capítulo apresentando sua natureza interdisciplinar, bem como sua relação específica com Ciência da Computação (CC), Ciências Cognitivas – Inteligência Artificial e Lingüística – Terminologia, que parecem ser as que justificam sua aproximação com a WS, no que diz respeito à RI, representação da informação e do conhecimento.

- Capítulo 3: discorre sobre a relação da tecnologia com a sociedade. Apresenta a tecnologia de forma geral como uma forma de viver da sociedade, principalmente em relação às TICs, que se apresentam como uma extensão da memória social, colaborando para a existência de uma “inteligência coletiva”. Apresenta os conceitos, origens e características da Internet e da WWW.
- Capítulo 4: objetiva apresentar três processos à luz da CI: recuperação da informação, representação da informação e representação do conhecimento, no que diz respeito aos seus aspectos conceituais e característicos, além de alguns instrumentos para aplicação desses processos. Capítulo 5: apresenta-se um panorama da WS, através de duas dialéticas: Web atual versus WS; WS início versus WS atualmente. A partir da seleção de alguns textos foi desenvolvido um panorama evolutivo da WS, composto principalmente, por conceitos, origens, propostas, procurando verificar o que foi ou está sendo desenvolvido.
- • Capítulo 6: a WS na CI, através das teses e dissertações de programas de pós-graduação da área no Brasil. Na primeira parte têm-se os elementos que caracterizam a identificação desses trabalhos e na segunda, um levantamento dos elementos que caracterizam seus conteúdos. Na última parte há a análise individual de cada trabalho, a partir de seus objetivos e de como a WS é usada.
- Capítulo 7: delineamento de algumas considerações e sugestões identificadas no trabalho.

2 CIÊNCIA DA INFORMAÇÃO

Esse capítulo descreve a CI, apresentando inicialmente seu contexto científico, caracterizado no presente trabalho como pós-moderno. Em seguida são indicados suas origens e seus períodos, a relação com a Biblioteconomia e a Documentação, disciplinas que colaboram para a origem da área, além da relação e dependência com as/das TICs. Os conceitos de CI aparecem a seguir, assim como o seu objeto, que também nomeia essa ciência, a informação. A natureza interdisciplinar da área encerra o capítulo, expondo especificamente as relações com a Ciência da Computação (CC), Ciências Cognitivas – Inteligência Artificial (IA) e a Linguística – Terminologia, que são importantes para o relacionamento entre a CI e a WS, justificando assim o trabalho.

2.1 O contexto científico da Ciência da informação

*E, de alguma forma, o paradigma dominante até hoje,
que começa a ficar um pouco combalido,
era um paradigma de disjunção e de redução.
Queremos conhecer separando, ou desunindo,
a ciência, a filosofia, a cultura literária,
a cultura científica, as disciplinas, a vida, a matéria, o homem, etc.
Edgar Morin*

A constituição da ciência talvez possa ser melhor explicada através de paradigmas do que historicamente (no que diz respeito a um processo linear ou cronológico). Não é negada que a análise histórica da ciência é também importante para sua explicação, mas parece claro que esse tipo de análise está voltado para acontecimentos, enquanto que uma análise paradigmática está voltada para o entendimento de idéias. Capurro (2003) define paradigma afirmando que se trata de um exemplar, de apresentar algo frente a uma referência ou a uma analogia. Assim surge um momento em que um paradigma apresenta seus limites e acontece uma ruptura. No caso das teorias científicas, ocorre uma revolução científica, e surge, portanto um novo paradigma. Kuhn (1996, p.13) define paradigma como “as realizações científicas universalmente reconhecidas que, durante algum tempo, fornecem problemas e soluções modelares para uma comunidade de praticantes de uma ciência”.

Sendo assim, recorrendo a Capurro (2003, p.3, grifo nosso), existe a necessidade de citar outros elementos que influenciam a constituição da ciência e de seus paradigmas:

[...] crises, rupturas, erros, mal entendidos, equívocos, analogias, dados empíricos, conceitos, hipóteses, dúvidas, retrocessos e buscas sem saída assim como as **instituições**, os instrumentos, as visões e paixões que suportam por assim dizer os processos cognitivos, constituem o cerne mesmo, em parte latente e em parte explícito, de todo campo científico [...]

O grifo na citação anterior nos leva a Stokes (2005) e a seu trabalho intitulado “O Quadrante de Pasteur”, em que ele destaca a institucionalização da ciência (moderna), podendo-se afirmar que as instituições exercem maior influência na constituição da ciência. Como toda e qualquer atividade da sociedade, a ciência necessita de investimentos, organização e regulação, ou seja, precisa se institucionalizar.

A ciência, como a entendemos hoje, surge com os gregos, “os inventores da investigação científica”, pois eles “estavam dispostos a encarar o mundo como um sistema natural governado por causas naturais gerais e identificáveis” (STOKES, 2005, p.52). Já Marcondes (2001)¹⁷ citado por Francelin (2004, p.50) afirma que a ciência como se conhece hoje, surge no século XVIII, “como tentativa de rigoroso exercício da razão”. Contudo Francelin (2004) não omite os gregos, afirmando que o período inicial da razão no pensamento do Ocidente se estabelece na Idade Antiga (Clássica), como a filosofia de pensadores, em um período que vai do século VI antes de Cristo e se encerra no século I da era cristã.

Nesse processo de sua constituição, a ciência passa por um período de estagnação, que acontece na Idade Média, entre os séculos V e XV, período em que a Igreja no Ocidente utiliza a doutrina cristã de forma a controlar o pensamento das pessoas. Um processo de transição ocorre no final desse período (séculos XV e XVI), dando início à Idade Moderna, onde as pessoas se afastam da crença e se aproximam (novamente) da razão. (FRANCELIN, 2004). Surge, portanto, a Ciência Moderna.

Segundo Guimarães e Silva (1999, p.81), com o modelo da racionalidade, a Ciência Moderna busca a formulação de leis, com idéias matemáticas de observação, usando o chamado método científico, para “dividir e classificar para determinar as relações do que foi separado”, período onde “conhecer é sinônimo de quantificar”, e onde o que “não é quantificável, é cientificamente irrelevante”. Período em que o método científico, fator fundamental na produção da ciência moderna torna o

¹⁷ MARCONDES, D. **Iniciação à história da filosofia**. 6. ed. Rio de Janeiro: Jorge Zahar, 2001.

“como fazer mais importante que o próprio sujeito, que o agente ou que os fins” (GUIMARÃES E SILVA, 1999, p.81). Período em que o homem, como dominador racional, deixa de lado o sujeito e se prende à forma de como fazer ciência. Francelin (2004, p.51) resume esse período, afirmando:

A Idade Moderna comporta as primeiras e substanciais mudanças no desenvolvimento do pensamento filosófico científico. Copérnico (1473-1543) defende o heliocentrismo em oposição ao teocentrismo ; Galileu (1564-1642) transporta o pensamento ao “experimento”; Bacon (1561-1626) aplica o método visando à objetividade científica, Descartes (1596-1650) desenvolve a metodologia em direção à verdade; e Newton (1642-1727) isola o objeto e entende o mundo como uma máquina que pode ser dividida em partes. A modernidade se caracteriza, portanto, pela supremacia do ser enquanto pensador e enquanto dominador do mundo pela razão.

Ainda sobre a Ciência Moderna, Guimarães e Silva (1999, p.84) identifica alguns pressupostos, a saber:

a) que há o domínio da ordem científica; b) que esta ordem separa os saberes; c) que a produção do conhecimento é fragmentada; d) que o saber é valorizado, distribuído e repartido em função dessa ordem; e) que a ordem apoia-se em um suporte institucional, que a legitima; f) que existem sistemas de exclusão dentro dessa ordem; g) que existem grupos que partilham dessa ordem, reforçando a permanência da situação através de um conjunto de práticas.

Morin (1999) aponta três pressupostos para a Ciência, nessa concepção moderna: ordem (determinismo universal), separação (exemplificado pela separação das disciplinas) e razão (coerência, autenticada pelos pressupostos anteriores). Tais princípios parecem “constituir o fundamento absolutamente incontestável do saber” (MORIN, 1999, p.23).

Essa ordem na Ciência Moderna como citado, pode ser exemplificada pela resposta de Morin (2003b, p.119) para a pergunta “o que é ciência?”. Para esse autor, não há resposta científica para tal questão, pois “a última descoberta da epistemologia anglo-saxônica afirma ser científico aquilo que é reconhecido como tal pela maioria dos cientistas”. Ou seja, a Ciência Moderna, numa tendência de estimular e valorizar fortemente seus postulados, deixa de legitimar outras visões do mundo, impedindo, segundo Alves (1994, p.155)¹⁸ citado por Guimarães e Silva (1999, p.84), que se vejam “as coisas tais como são vistas por outros que utilizam padrões diferentes”.

¹⁸ ALVES, R. **Filosofia da ciência**: introdução ao jogo e suas regras. 20.ed. São Paulo: Brasiliense, 1994.

Esse paradigma moderno na constituição da ciência predominou fortemente através da institucionalização, como bem apresenta Stokes (2005). Ou seja, é extremamente difícil quebrar essas regras de construção do conhecimento chamado científico, uma vez que as instituições que recebem e financiam a ciência realizam dessa forma a geração de conhecimento, dividindo, reduzindo, negando outras formas de saber. Isso era evidenciado pelo afastamento total do senso comum, que se entende aqui como a realidade cotidiana, ou seja, o que é usado para viver (FRANCELIN, 2004; GUIMARÃES e SILVA, 1999).

Nesse sentido, Stokes (2005) propõe seus quatro quadrantes: pesquisa básica pura (Bohr); pesquisa aplicada pura (Edison); pesquisa de fenômenos particulares sem objetivos gerais; e a pesquisa básica inspirada pelo uso (o Quadrante de Pasteur, que dá título ao seu livro). O autor indica o Quadrante de Pasteur para demonstrar como o paradigma vigente dificultava o desenvolvimento da ciência, uma vez que a ordem da ciência estabelecida pelo paradigma moderno separava o entendimento e aplicação da ciência, além de desconsiderar o conhecimento gerado pela aplicação (técnica) (STOKES, 2005).

Evidentemente, a Ciência Moderna tem suas contribuições, mas “o mito da ciência que tudo explica e que a tudo domina, tendo por fundamento uma razão que se considera única e universal, começa a sofrer críticas dentro da própria modernidade”, como afirma Francelin (2004, p.51). Morin (2003b, p.118) também afirma que “o exemplo de Einstein é elucidativo”, pois “o espírito mais genial não dispõe de condições que lhe permitam pensar a ciência na sociedade, isto é, conhecer o lugar e o papel da ciência na sociedade”. Ou seja, parece que a ciência, nessa concepção moderna já não responde a tudo, Schwartz (1992, p.17)¹⁹ citado por Guimarães e Silva (1999, p.90) afirma o seguinte:

Quase tudo o que se lê hoje em dia sobre ciência está, no mínimo, incompleto, uma vez que geralmente considera-se uma parte do todo, não observando-se o contexto em que se trabalha. Os cientistas estão em busca da verdade? Não. Seria correto dizer que eles procuram entendimento. A ciência é uma questão de objetividade? Não. É uma questão de provas, argumentação, persuasão e, acima de tudo, de poder. A ciência é feita de números? Somente em parte. Fundamentalmente, ela é feita de relações. A ciência é um conjunto de narrativas escritas a respeito do relacionamento com a natureza que se acumulou de geração em geração.

¹⁹ SCHWARTZ, J. **O momento criativo: mito e alienação na ciência moderna**. São Paulo: Best-Seller, 1992.

Surge, portanto, a ciência numa perspectiva Pós-Moderna (século XX), caracterizada pela incerteza e pela desordem (MORIN, 1999, p.24), pois “é apenas em casos muito privilegiados, em situações especiais, que nos encontramos verdadeiramente confrontados com variáveis exatas cuja definição é clara e sem ambigüidade” diz Moles (1995, p.15)²⁰ citado por Guimarães e Silva (1999, p.90).

Francelin (2004, p.51) afirma o seguinte sobre a pós-modernidade:

A pós-modernidade é marcada por profundas transformações em diversos setores do conhecimento e do próprio cotidiano humano. Os avanços industriais, as guerras, o comércio, a produção em massa, a divisão política e econômica dos países, assim como o impulso tecnológico, interferem e sofrem interferência quase que direta do pensar e do fazer ciência. Estas questões fazem com que o ser humano se obrigue a ter posturas diferenciadas, ou seja, cada vez mais o mundo se pluraliza. A pluralização, por sua vez, também exige um pensamento plural.

Mas apesar do reconhecimento dessa perspectiva pós-moderna da ciência, vale retomar o conceito de paradigma proposto inicialmente, salientando que o modelo chamado moderno da ciência ainda existe e predomina, ou seja, passa-se por um período de transição e coexistência dessas duas idéias, apesar de uma evidente ruptura.

A pós-modernidade é um termo recente, possuindo entre trinta e cinquenta anos, mas com uma idéia que se pode chamar de centenária, que influenciou não apenas a ciência, mas também outras manifestações do homem na sociedade, segundo Francelin (2004) citando diversos autores.

Na ciência “a fragmentação pós-moderna não é disciplinar e sim temática, na medida em que a noção de interdisciplinaridade relaciona-se à construção de um projeto comum”, portanto, a construção “de um produto através de proposições oriundas de vários campos do saber”. (ORRICO, 1999, p.147). Ou seja, a ciência nesse contexto se pauta pela noção de finalidade, como apontam Orrico (1999) e Stokes (2005).

A pós-modernidade é, portanto, um período de flexibilidade, um período plural e não de fragmentos (FRANCELIN, 2004). Isto é, não se trata de um paradigma científico de incerteza e desordem, mas sim de busca de uma maneira diferente para lidar com esses elementos presentes no desenvolver do conhecimento, numa dialética com o paradigma moderno. Um exemplo são as Ciências Sociais, que surgem no século XIX, “que não tem – ou quase não tem – ciência e métodos que lhe sejam

²⁰ MOLES, A. **As ciências do impreciso**. Rio de Janeiro: Civilização Brasileira, 1995.

próprios” (MOLES, 1995, p.17²¹ citado por GUIMARÃES e SILVA, 1999, p.91).

Santos (1993) aponta alguns obstáculos na busca das Ciências Sociais pelo estatuto científico, como: fenômenos sociais estudados como naturais; fatos sociais reduzidos a coisas; leis universais dificilmente estabelecidas porque o objeto social é historicamente condicionado e culturalmente determinado; fenômenos sociais são subjetivos, dificultando as previsões (o indivíduo altera o seu comportamento na medida em que adquire conhecimento sobre ele e o seu ambiente); cientistas sociais não se despem dos valores que orientam sua prática, quando observam um fenômeno (isenção e neutralidade).

Vale, portanto, resgatar dois paradigmas que se contextualizam na pós-modernidade: a Complexidade e o Paradigma Emergente. Pelo primeiro entende-se que muitos postulados do paradigma moderno da ciência caem por terra, não há mais certeza nos experimentos científicos, descobrem-se coisas que não podem ser explicadas, que o mundo pode ser um sistema, onde não se pode entender o todo sem as partes, nem essas últimas sem o todo, como propõe Francelin (2003), citando autores como Morin, Moles e Von Bertalanffy. Morin (1999, p.28) apresentando o que chama de complexidade, afirma:

O sistema, como já foi dito – o todo –, é mais que a soma das partes, isto é, no nível do todo organizado há emergências e qualidades que não existem nas partes quando são isoladas. Tais emergências podem retroagir sobre as partes: a cultura é uma emergência social que retroage sobre os indivíduos, lhes dá a linguagem e o saber, e, por isso, os transforma.

O Paradigma Emergente é proposto por Boaventura Santos e é assim chamado, pois o autor entende que há um outro dominante, que reflete os postulados do paradigma moderno da ciência. Não difere muito da Complexidade proposta por Morin, pois trata-se da soma do paradigma científico com o social, onde: o conhecimento científico-natural é conhecimento social, visando o senso comum e auto-conhecimento; o texto científico é transdisciplinar e individualizado; há pluralidade na metodologia com migração de teorias e conceitos, criação de contextos e da situação comunicativa; ou seja, o conhecimento é local e total. (GUIMARÃES e SILVA, 1999; ORRICO, 1999; SANTOS, 1993).

A ciência nessa perspectiva pós-moderna busca uma aproximação com o senso comum, em um sentido de troca, influenciando e sendo influenciada, elaborando suas investigações pelas temáticas

²¹ Idem 20.

e não mais somente pelas disciplinas, pois desta forma parece que sua contribuição para a sociedade será melhor.

Nesse cenário surge a CI, que se caracteriza por postulados que não se encaixam no contexto moderno de ciência. Isso pode ser exemplificado por Silva (2005), que reconhece tal situação na pesquisa em CI no Brasil, apontando a busca para consolidar conceitos fundamentais que alicerces a área. Le Coadic (1996, p.22) reforça essa idéia, afirmando que “a teoria, portanto, apresenta-se atrasada em comparação com o empírico e existe, sobretudo, uma falta de ligação entre os dois”.

Pela presença do empirismo na CI, pode-se deduzir que a CI é uma ciência social. Le Coadic (1996) indica esse caráter social da CI, afirmando que sua preocupação está em um problema social, a informação. No Brasil, Araújo (2003) reforça esta idéia, indicando que, para agências de fomento de pesquisa e algumas Universidades a classificação da CI é de uma Ciência Social Aplicada. Mas, apesar de nascer em um período “em que já se observam as primeiras críticas ao fracasso do projeto da modernidade e ao modelo científico resultante desse projeto” (ARAÚJO, 2003, p.22), a CI parece ainda buscar parâmetros no modelo moderno da ciência, procurando leis universais no comportamento da informação.

Isso pode ser justificado pelo presente sistema de institucionalização da ciência, que ainda predomina conforme afirma Stokes (2005). No entanto, através de estudos de Francelin (2003; 2004), Guimarães e Silva (1999) e Orrico (1999), parece correto afirmar que o contexto científico da CI se enquadra numa perspectiva Pós-Moderna, com abordagens como a da Complexidade de Edgar Morin e a do Paradigma Emergente de Boaventura Santos.

Para justificar tal afirmação se torna necessário enunciar algumas características da CI. A primeira é sua origem no século XX, época em que se consolida a idéia de uma ciência que questiona os postulados da ciência moderna. Outras características da CI que dificultam a aceitação de sua cientificidade são: a não descoberta de novas leis, nem invenção de novas teorias, segundo Machlup e Mansfield (1983, p.12)²² citados por Guimarães e Silva (1999, p.103); manifestações da forma de ser da ciência pós-moderna que possui percepção difícil, uma vez que elas não estão institucionalizadas ou “esquematizadas” dentro dos padrões esperados (CHRISTOVÃO, 1995, p.33)²³ citado por GUIMARÃES e SILVA, 1999); disciplina que não se baliza pelo modelo clássico e não possui teoria, mas volta suas pesquisas para a solução de problemas e que não possuirá uma

²² MACHLUP, F.; MANSFIELD, U. **The study of information**: interdisciplinary messages. New York: John Wiley and Sons, 1983.

²³ CHRISTOVÃO, H. T. A Ciência da Informação no contexto da pós-graduação do IBICT. **Informare**: cadernos de Programa da Pós-Graduação em Ciência da Informação, Rio de Janeiro, v.1, n.1, p.25-30, jan./jun. 1995.

teoria, mas conceitos amplos e reformulados. (WERSIG, 1993); bases multidisciplinares e tratáveis somente com a união dos campos envolvidos, na busca por uma síntese, diz Heilprin (1989, p.343)²⁴ citado por Guimarães e Silva (1999); característica interdisciplinar, como mostram Borko (1968), Freire (2006), Pinheiro (1999) e Saracevic (1996).

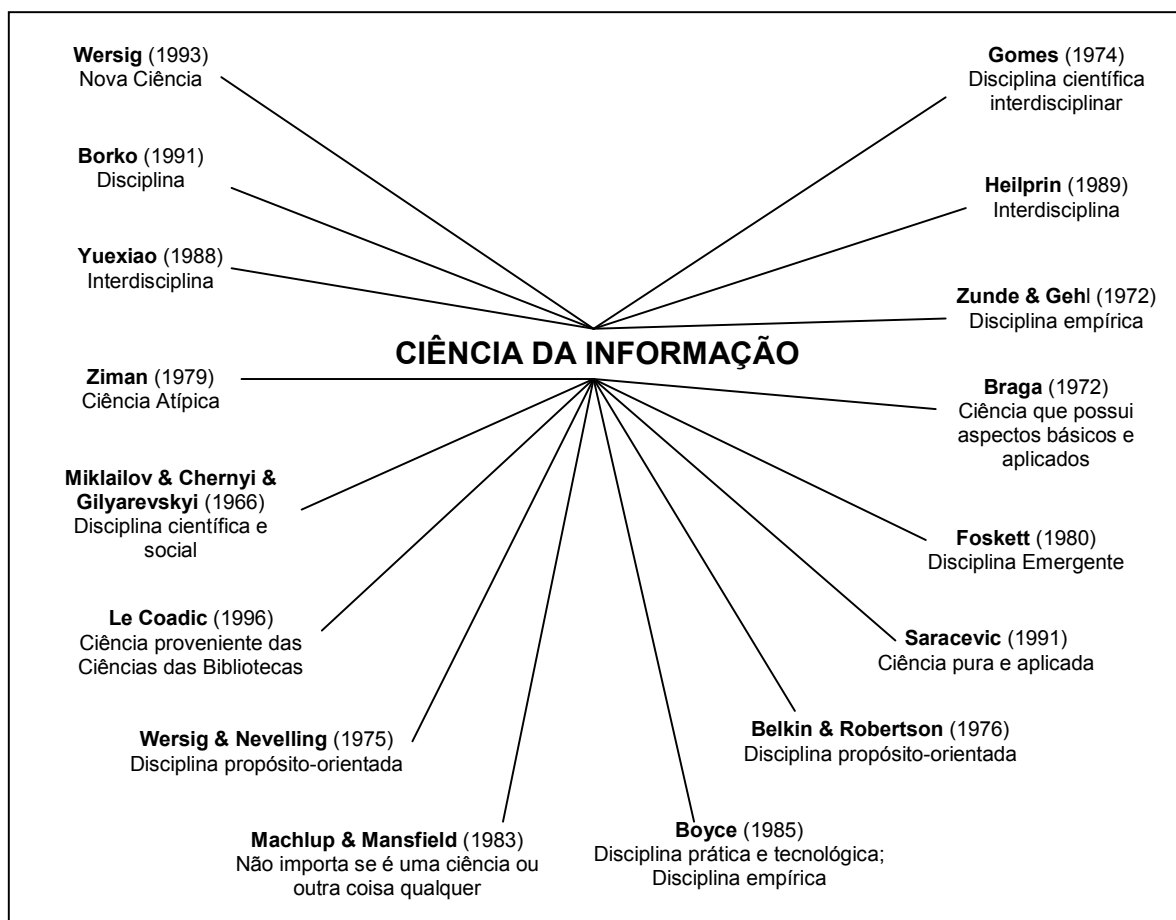
Esse conjunto de obstáculos que contribui para a aceitação da cientificidade da CI, segundo Machlup e Mansfield (1983, p.12)²⁵ citados por Guimarães e Silva (1999, p.104) “não reduz a sua significância e não deveria reduzir a sua respeitabilidade científica”. Os estudiosos da ciência que têm discutido os limites das diversas disciplinas, “não desenharam mapas mostrando como certos fenômenos, problemas, leis, conceitos e teorias podem ser apropriadamente transferidos para várias disciplinas”. Assim, podemos afirmar que desde que haja pesquisa para as soluções dos diversos problemas informacionais na sociedade, há um espaço para a CI, como uma disciplina científica não do parâmetro moderno, mas certamente uma disciplina científica que agrega saberes de várias outras áreas do conhecimento com um foco próprio, pois trabalha diante de problemas que em certa medida são apenas seus, como os estudos bibliométricos, as representações da informação e do conhecimento, pautadas pela necessidade de entender e representar os documentos produzidos para recuperação.

Guimarães e Silva (1999) apontam a diversidade de percepções sobre a CI, através da figura a seguir.

²⁴ HEILPRIN, L. B. Foundation of Information Science reexamined. **Annual Review of Information Science and Technology** (ARIST)., White Plains – NY, v.24, p.343-372, 1989.

²⁵ Idem 22.

Figura 2. Percepções sobre a Ciência da Informação enquanto disciplina científica



Fonte: GUIMARÃES e SILVA, 1999, p.105.

Com base no exposto, parece possível dizer que a CI é uma disciplina científica do paradigma pós-moderno, uma disciplina temática, que cuida da informação e da sua influência na sociedade, que deve fazer uso dos vários procedimentos, técnicas, conceitos e métodos de outras disciplinas para validar seus postulados. Podemos concluir, portanto, que as atividades, sejam elas científicas ou tecnológicas, que beneficiam a sociedade podem e devem se valer de tudo o que for possível para atingir seus objetivos, pois na verdade toda ciência deve apresentar elementos que interajam com a sociedade, ou seja, focar o homem.

2.2 Origem e períodos da Ciência da Informação

A origem oficial da CI é estabelecida por alguns autores após a Segunda Guerra Mundial (1939-1945), período em que ocorre uma produção científica muito grande, gerando muita informação registrada (BARRETO, 2002; MIRANDA, 2003; SARACEVIC, 1996), através de relações com a Biblioteconomia, a Documentação e as TICs, principalmente.

Considera-se o artigo do americano Vannevar Bush como um dos marcos no início dessa nova ciência. Esse autor questiona a forma artificial de organização da informação presente naquele período e que ainda persiste, afirmando que ela não responde aos interesses de quem busca informação (BUSH, 1945). Há, portanto a necessidade de mudar a forma de organização do material bibliográfico produzido, e nesse sentido Bush (1945) afirma que ela deve adotar como base a maneira como os indivíduos pensam, ou seja, por associação de conceitos.

A CI em sua origem recebeu outra influência, a européia, e o marco dessa influência está na obra de Paul Otlet: “*Traité de documentation*” (Tratado de documentação) de 1934. A proposta da Documentação é do final do século XIX. A questão da época era a criação de uma bibliografia universal. A Documentação influenciou fortemente o surgimento da CI (ORTEGA, 2002). Na sua definição de Documentação, Smit (1986, p.10) afirma que enquanto a biblioteca organiza seus documentos focando o acervo e funcionando em função dele, “a documentação organiza as informações relacionadas a um assunto, sem restrições quanto ao acervo [...]”, argumento reforçado por Freire (2006).

A Biblioteconomia influenciou a criação da Documentação, já que a CDU²⁶ foi criada a partir da CDD²⁷ (SHERA, 1980), mas também influenciou a CI, uma vez que a proposta dessa nova ciência questionou métodos do tratamento da informação da Biblioteconomia, ou melhor, começou a propor outros métodos, através do uso das tecnologias. Goffman (1970)²⁸ citado por Mostafa (1985, p.21) diz que a CI “precisa transcender as bibliotecas e os computadores e desenvolver seus princípios independentemente desses ou quaisquer outros sistemas físicos”.

A origem da CI não é unanimidade na literatura da área. A exposição dos períodos da área por Barreto (2002) e Robredo (2003) indicam posições divergentes. O primeiro propõe três períodos, a

²⁶ Classificação Decimal Universal – Instrumento para a classificação de materiais bibliográficos, utilizado para a organização de bibliotecas, idealizado por Paul Otlet e Henry La Fontaine. Sua primeira edição foi publicada entre 1904 e 1907.

²⁷ Classificação Decimal de Dewey – Instrumento utilizado para classificação e organização de material bibliográfico nas estantes de bibliotecas, criado por Melvil Dewey, publicado pela primeira vez em 1876 (ROBREDO, 2005).

²⁸ GOFFMAN, W. Information science: discipline or disappearance. *ASLIB Proceedings*, London, v. 22 n.12, p.589-596, 1970.

saber: “tempo da gerência de informação que vai de 1945 a 1980; tempo da relação informação e conhecimento no período de 1980 a 1995; tempo do conhecimento interativo de 1995 até os dias atuais” (BARRETO, 2002, p.71).

Robredo (2003) apresenta os seguintes períodos: dos pioneiros da documentação à Guerra Mundial (período da concepção e fundamentos da documentação); de 1919 à Segunda Guerra Mundial em 1945 (período do estabelecimento do Instituto Internacional de Documentação, da publicação do “Tratado de Documentação” de Paul Otlet e do movimento de bibliotecas especializadas tanto na Europa como nos Estados Unidos); do pós-guerra à Conferência Internacional sobre Informação Científica, em 1958 (período da explosão documental e comprovada ineficiência para a organização dos documentos); de 1960 à década de 80 (período das discussões e propostas quanto ao campo científico da CI e suas primeiras definições); da década de 90 aos dias atuais (período de aprofundamento da área como campo científico, de discussão dos vários conceitos que a envolvem, e também de sua natureza e da relação interdisciplinar).

Pode-se perceber a partir dessas propostas, as divergências para o surgimento da CI. Mas, a indicação de Barreto (2002) é mais temática, enquanto a de Robredo (2003) parece ser histórica, ou seja, vale considerar ambas descrições do surgimento da área, mesmo por que elas não se excluem. Assim, relações com a Biblioteconomia e a Documentação, além da relação com as TICs colaboraram para o surgimento da CI.

2.2.1 Biblioteconomia

Segundo Le Coadic (1996), a Biblioteconomia é uma prática de organização, sendo uma das disciplinas da CI, que responde por acervos, por administração da biblioteca enquanto organização e por seus usuários. Sua história se estabelece ao longo dos últimos três mil anos. O termo foi utilizado pela primeira vez em 1839 em uma obra francesa intitulada “Bibliothéconomie: instructions sur l’arrangement, la conservation e l’administration des bibliothèques”, mas foi Gabriel Naudé, em 1627 que estabeleceu os primeiros princípios da Biblioteconomia chamada moderna

(LAHARY, 1997)²⁹ citado por (ORTEGA, 2002).

Shera (1972)³⁰ citado por Saracevic (1996) diz que as bibliotecas contribuem para o sistema de comunicação na sociedade, aumentando a utilização dos registros para benefício dos indivíduos. Esse é o aspecto em que ela se assemelha à CI, segundo Saracevic (1996), ainda mesmo que havendo diferenças e relações entre ambas, o autor as considera como campos distintos. Reforçando essa idéia, há quem entenda biblioteconomia como uma ciência social, numa perspectiva contemporânea.

A tendência fundamental da biblioteconomia contemporânea é a concentração crescente no estudo dos aspectos sociais, na análise da função econômica, social e cultural das bibliotecas; o exame das relações mútuas entre o livro e sociedade é que lhe dá a característica de uma ciência social. (BRANDÃO, 1982, p.24).

Parece correto afirmar que a Biblioteconomia tem um grande papel ao trabalhar com a informação, devido principalmente ao espaço institucional e também aos seus profissionais, os bibliotecários. Isso acontece no contexto americano, europeu e brasileiro. Portanto, devido a essa aproximação com a Biblioteconomia, a CI se beneficia desses aspectos colocados acima. Esse trabalho entende a Biblioteconomia como uma área de atuação da CI, apesar da primeira existir antes da segunda, ou mesmo ser uma disciplina da CI como colocado por Le Coadic (1996).

2.2.2 Documentação

A Documentação talvez seja o campo que mais se identifica com a CI. Smit (1986) chega a afirmar que muitos chamavam a Documentação de CI. Ortega (2002) apresenta autores que defendem os primórdios da Documentação no século XV, com a construção de bibliografias universais. Mas entendemos que a Documentação surge no final do século XIX, quando os pesquisadores diante da necessidade de explorar um número cada vez maior de documentos variados sentiram a necessidade de novos esquemas para organizar os documentos, uma vez que a Biblioteconomia já não respondia

²⁹ LAHARY, D. *Bibliothéconomie*. In: CACALY, S. (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation**. Paris: Nathan, 1997. p.78-81.

³⁰ SHERA, J.H. **The foundations of education for librarianship**. New York: Becker and Hayes, 1972.

a essa demanda. (LE COADIC, 1996).

O pioneiro da Documentação foi o belga Paul Otlet, que em 1892, encontra-se com Henri La Fontaine e em 1895, ambos criam o Instituto Internacional de Bibliografia (IIB), com o objetivo de estabelecer uma bibliografia universal (ORTEGA, 2002; RAYWARD, 1997).

Uma das propostas da Documentação está no conceito do que é um documento.

Documentos consistem do que quer que represente ou expresse um objeto, um fato, uma impressão por meio de qualquer tipo de sinal (escrita, pintura, diagramas, símbolos). [...] documentos escritos (livros, periódicos, jornais e manuscritos), as ilustrações (gravuras, desenhos, fotografias), os ideogramas (mapas, plantas, quadros, sinópticos, diagramas) e a música. (OTLET, 1990³¹, p.105 citado por RAYWARD, 1997).

Shera (1980) recorre a Briet³², para afirmar que até um animal catalogado em um zoológico é um documento, pois se trata de algo que informa. O mais importante na Documentação é o conteúdo e a finalidade do serviço, não o tipo de organização. (BRANDÃO, 1982).

Nesse sentido, a documentação, segundo Otlet (1934)³³ citado por Shera (1980), trata de uma série de técnicas que se prestam a reunir, organizar e difundir a informação contida em qualquer tipo de documento. Há várias definições sobre o que seja a documentação. Shera (1980) propõe a sua, enfatizando que se trata de um serviço que cuida da circulação da informação, em nível de especialidade e não em nível popular, no sentido do que o autor entende ser a Biblioteconomia Especializada.

Para entender melhor os objetivos da Documentação, Rayward (1997) lista seis atividades apresentadas por Otlet: 1) produção de trabalhos com descobertas; 2) coleção desses trabalhos; 3) análise do trabalho para a elaboração de resumo; 4) redistribuição de trabalhos similares; 5) codificação; e 6) enciclopédia.

Todas essas etapas devem atender a uma estrutura que supra a necessidade do usuário da informação.

³¹ OTLET, P. The systematic organisation of documentation and the development of the International Institute of Bibliography. Institut International de Bibliography Publication No. 82, 1907. In: RAYWARD, W. B. (Ed. and Trans.). **The international organization and dissemination of knowledge**: selected essays of Paul Otlet. Amsterdam: Elsevier, 1990.

³² Suzanne Renée Briet: bibliotecária francesa que teve um papel fundamental no movimento da Documentação: criou uma escola para essa disciplina e ficou conhecida como a Madame Documentação (BUCKLAND, 2005).

³³ OTLET, P. **Traité de documentation: le livre sur le livre**: théorie et pratique. Bruxelles: Mundaneum, 1934.

A partir das afirmações expostas, podemos entender que a Documentação apresenta vários postulados que estão presentes na CI: foco no usuário, finalidade mais importante que o processo no tratamento da informação; informação organizada a partir dos conteúdos; informação atualizada disponível o mais rápido possível.

Pode-se afirmar, portanto, que a Documentação foi o primórdio do que entendemos por CI. No entanto, na época de sua criação, não se contava com as tecnologias, principalmente as de informação e comunicação, que favoreceram o surgimento da CI.

2.2.3 Tecnologias de informação e comunicação

Quando pensamos em tecnologia atualmente, podemos pensar em eletrodomésticos, carros, Internet, Web, computadores, ou seja, uma série de artefatos que podem realizar atividades que auxiliam as pessoas. Trata-se de algo que predomina na sociedade atual (anos 1950 em diante), pois parece correto afirmar que, se estamos na Sociedade da Informação, isso se deve às tecnologias. Essa parte do trabalho apresenta uma breve relação das TICs, de forma geral com a CI, pois segundo alguns autores esse tipo de tecnologia é indispensável quando se fala em origem e desenvolvimento dessa área do conhecimento.

As tecnologias modificaram o modo como o homem vive, pensa e se relaciona, pois exercem influência social, principalmente, as “tecnologias da informática”, ou seja, as TICs. Lévy (1998, p.107) afirma que:

Os computadores permitem utilizar sofisticadas ferramentas estatísticas, dão acesso à simulação de modelos digitais e contribuem amplamente para a formalização dos procedimentos da pesquisa. Assim podem conferir um inédito caráter de rigor e precisão às ciências humanas. A informática seria, portanto, mais do que um instrumento científico. Capaz de fazer uma disciplina passar de um estado pré-científico ao estatuto de ciência exata, seria um operador metacientífico. No atual debate epistemológico, o computador reforça o campo do método científico único, pois pela primeira vez proporciona-lhe os meios de suas ambições.

Temos, portanto uma influência das TICs em todos os setores da sociedade, pois, parece difícil pensar em medicina, economia, estatística, arquitetura, administração, educação, e até mesmo filosofia, além de informação e de comunicação sem o uso de artefatos tecnológicos.

Para Targino (1995, p.3):

A influência da informática nos processos de difusão da informação é decisiva. Conduz à adoção de novo modelo informacional distributivo, dinâmico e hipertextual, no sentido de atender os usuários conectados em redes eletrônicas, de modo não mais linear, mas respeitando sua estrutura cognitiva, suas demandas singulares, independentemente da localização geográfica. Tal dinamicidade pressupõe maior agilidade, precisão, completeza, consistência e densidade.

Essa necessidade em relações as Tics foi comentada por Bush (1945), abrangendo a forma artificial adotada para a organização informação, com base em uma teoria inapropriada para as necessidades presentes no pós-guerra, reflexo de uma formação profissional defasada e o instrumental (tecnológico) de armazenamento e de recuperação para essa informação; ou seja, existiu a necessidade de tecnologias apropriadas para um efetivo tratamento da informação que foi sendo gerada. Mas apenas nos anos de 1980 o computador tornou-se acessível à CI (BARRETO, 2002).

A emergência de uma sociedade em transição, onde a informação é um bem com valor agregado e as empresas que mais têm lucros são as ligadas à ela (Microsoft, Google, Apple, etc.), evidenciam que a informação virou um produto, um bem comercial, um fator econômico, e junto com as novas tecnologias permitiu a criação de novos mercados, serviços, empregos e empresas (BORGES, 2000).

Uma das primeiras propostas para “mecanização” de atividades ligadas à informação foi também de Bush (1945), com o “MEMEX” (“*memory extension*”, ou “*extensão da memória*”) um aparelho tecnológico que armazenava e recuperava documentos mediante a associação de palavras (BARRETO, 2002).

Podemos entender melhor a finalidade das TIC’s para a CI através da seguinte afirmação, que apresenta a “informática documentária”, como:

O conjunto de aplicações da Informática Documentária, técnica que se refere às intervenções da informática nas diversas fases de produção e utilização de documentos: produção de textos, difusão pelo editor, gestão da biblioteca, análise e indexação para a constituição de bases de dados bibliográficos e para a difusão seletiva, e softwares para a pesquisa nestas bases de dados. (DEWEZE³⁴, 1994, p.1 citado por ORTEGA, 2002, p.29).

³⁴ DEWEZE, A. **Informatique documentaire**. 4.ed. Paris: Masson, 1994. 292p. (Manuels informatiques Masson)

Barreto (2002, p.73) reforça essa relação, afirmando que “são as novas tecnologias de informação e comunicação que modificaram aspectos fundamentais, tanto da condição da informação quanto da condição da comunicação”, modificando enfatiza o autor “radicalmente a qualidade de tempo e espaço entre as relações do emissor, os estoques e os receptores da informação”.

Diante do paradigma da computação em rede, onde o compartilhamento de recursos para redes de comunicação é fundamental (MEIRELLES, 1994), parece correto afirmar que se estreitou ainda mais a relação entre CI e TICs. A informação tem que estar em rede, tratada de forma a atender qualquer indivíduo que tenha acesso a ela. Ou seja, não da forma centralizada, como na bibliografia universal, idealizado por Otlet, mas de forma distribuída temos acesso a toda informação, sem levar em consideração espaço e tempo. Portanto, os desafios da CI continuam presentes, e ainda parecem maiores, pois o ambiente da Internet e da Web são ambientes em que todos (literalmente) podem influenciar.

2.3 Conceitos de Ciência da Informação

Assim como seu contexto científico, seu objeto de estudo, sua origem e suas relações, a CI possui ainda conceitos incertos, inacabados e em pleno processo de discussão, talvez principalmente por sua característica interdisciplinar, contando com pesquisadores originários de outros campos do conhecimento, influenciados por suas origens, caracterizando a CI com aspectos de várias áreas do conhecimento, sem uma profunda discussão.

Le Coadic (1996, p.26) afirma que a CI tem por objetivo o estudo das propriedades da informação (natureza, gênese, efeitos), “a análise dos processos de construção, comunicação e uso da informação; e a concepção dos produtos e sistemas que permitem sua construção, comunicação, armazenamento e uso”.

Segundo Borko (1968, p.3), a CI é “a ciência que investiga as propriedades e comportamento da informação, as forças que governam o fluxo da informação, e seus meios de processamento para o acesso e utilização máxima”. Os processos utilizados são: “origem, disseminação, coleção, organização, armazenagem, recuperação, interpretação e uso da informação”.

Para Shera (1980), o termo CI foi empregado para a Biblioteconomia não tradicional. A definição da

Conferência da Geórgia considera como atividades de um especialista em CI: o estudo e o desenvolvimento do armazenamento e da recuperação da informação, com propostas de novos métodos aplicáveis ao problema da informação.

Rees e Saracevic em uma Conferência em 1967 classificam a CI como um “ramo da pesquisa que toma substância, seus métodos e suas técnicas de diversas disciplinas para chegar a compreensão das propriedades, comportamento e circulação da informação”, incluem uma série de outras idéias para delineamento da área, como “análise de sistemas, os aspectos ecológicos da informação e da comunicação, dos meios de informação e da análise lingüística, da organização da informação, das relações homem/sistema, etc.”, por fim definem a CI como “estudo dos fenômenos da comunicação e das propriedades dos sistemas de comunicação” (REES; SARACEVIC, 1967³⁵ citados por SHERA, 1980, p.98).

Fazendo uma síntese a partir das definições apresentadas tem-se uma ciência que estuda e desenvolve esquemas ou sistemas de informação, usando tudo o que for possível para fazer com que essa informação circule e gere conhecimento novo, deixando de lado os estoques e considerando o fluxo da informação. Para Goffman (1970)³⁶ citado por Saracevic (1996, p.46), o objetivo da CI:

[...] deve ser o de estabelecer um enfoque científico homogêneo para estudo dos vários fenômenos que cercam a noção de informação, sejam eles encontrados nos processos biológicos, na existência humana ou nas máquinas... Conseqüentemente, o assunto deve estar ligado ao estabelecimento de um conjunto de princípios fundamentais que direcionam o comportamento em todo processo de comunicação e seus sistemas de informação associados... (A tarefa da CI) é o estudo das propriedades dos processos de comunicação que devem ser traduzidos no desenho de um sistema de informação apropriado para uma dada situação física.

Saracevic (1996, p.47) apresenta sua (re)definição da CI:

A CIÊNCIA DA INFORMAÇÃO é um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais.

³⁵ REES, A.; SARACEVIC, T. Education for Information Science and its relation to librarianship. Unpublished paper presented before the **Annual Conference of the Special Libraries Association**. New York, 1967.

³⁶ Idem 28.

Tem-se ainda a definição simples e objetiva de Belkin (1978)³⁷ citado por Ingwersen (1992) que afirma que a CI busca “facilitar a efetiva comunicação da informação desejada (armazenada) entre o gerador humano e o usuário humano”; nesse sentido apresentam-se cinco áreas de preocupação da CI: 1) transferência da informação em sistemas de comunicação cognitivos e humanos; 2) a idéia de informação desejada; 3) a efetividade de sistemas de transferência de informação; 4) a relação entre informação e gerador; e 5) informação e usuário.

Com base nas áreas apresentadas por Belkin, Ingwersen (1992, p.302) apresenta quatro importantes divisões da CI:

Bibliometria, isto é, o estudo quantitativo dos processos da comunicação escrita, como a co-citação, está principalmente interessada nas áreas 1 e 4. Administração da Informação, incluindo avaliação e qualidade dos sistemas de informação textual, está basicamente interessada nas áreas 3, 4 e 5. Desenho de sistemas de recuperação de informação pertence às áreas 2, 3 e 4. Recuperação de Informação (RI) está fundamentalmente preocupada com o estudo dos processos de informação nas áreas 2, 4 e 5.

Parece correto afirmar que as definições citadas estreitam ainda mais a dependência da CI em relação às TICs, como ferramentas, reforçando a questão cognitiva e o papel do homem como idealizador dos processos realizados pelas máquinas. O surgimento das tecnologias realmente impulsionou o aparecimento da CI, possibilitando que atividades impossíveis sem tais recursos acontecessem.

2.4 Objeto: a informação

Depois de contextualizar a CI como campo científico, é importante abordar seu principal objeto de estudo: a informação. Inicialmente vale salientar que a informação possui muitos conceitos. Nesse sentido, procura-se entender esse termo à luz da CI e sua importância para a sociedade.

Há quem entenda a informação como processo (SILVA, 2006), coisa (BUCKLAND, 1991) ou fenômeno (BARRETO, 1994; 1996). Le Coadic (1996, p.5) define a informação para a CI como um

³⁷ BELKIN, N. Information concepts for information science. *Journal of Documentation*, London, v.34, p.55-85. 1978.

“conhecimento inscrito sob a forma escrita, oral ou audiovisual”, que possui elemento de sentido. Esse caráter apresentado para o autor sobre a informação, a caracteriza como um objeto “tangível, mensurável, deteriorado física e intrinsecamente, com volume, peso, preço e outras propriedades administráveis”. (MIRANDA, 2003, p.177).

Barreto (1996, p.2) define a informação como “estruturas significantes com a competência de gerar conhecimento no indivíduo, em seu grupo ou a sociedade”. Nesse sentido temos a informação caracterizada pela sua finalidade, onde há uma aproximação necessária com o conhecimento, pois o mesmo autor afirma que o “conhecimento é toda a alteração provocada no estado cognitivo do indivíduo”, numa interação com uma estrutura de informação. (BARRETO, 1996, p.2).

A partir dessas duas primeiras definições apresentadas podemos afirmar que: a informação é algo que existe antes do conhecimento, um estoque de conhecimento inscrito em algum suporte; e que o conhecimento é um processo cognitivo que acontece a partir da consulta por um indivíduo a esse estoque. Popper (1994, p.68)³⁸ citado por Miranda (2003, p.175) expõe essa idéia afirmando que:

Existe uma diferença importante quando exprimimos o nosso pensamento por palavras ou, melhor ainda, por escrito. Desta forma o pensamento se torna acessível à crítica. [...]. Portanto, há, pelo menos, um significado importante de conhecimento objetivo: conhecimento no sentido de teoria formulada verbalmente, que se expõe à crítica. A isso chamo eu conhecimento no sentido objetivo. Nele se integram os conhecimentos científicos.

Silva (2006, p.195), afirma que a “informação” não é uma evolução de “dado”, nem o “conhecimento” uma evolução da “informação”, mas se trata de um “processo no qual o dado e o conhecimento unem-se em representação, em tradução da presença humana [...]”. Esse autor afirma ainda que “se considerarmos o ciberespaço, a informação não é estrutura estabelecida, não é resultado, não é produto, é todo um processo que acaba por nos tornar mais conscientes de que podemos adentrar outro processo, o de conhecer, e mais outro, o de agir” (SILVA, 2006, p.199).

A definição apresentada leva em consideração o espaço estabelecido pela Internet e pela Web, e, portanto, adequada, uma vez que estamos em uma sociedade em rede, principalmente devido às TICs. Vale salientar que os focos das definições são os indivíduos, ou melhor, transformações que venham a ocorrer nos indivíduos em contato com a informação, que pode melhorar o seu meio e conseqüentemente a sociedade. Portanto a informação se apresenta como elemento social fundamental, que “sintoniza o mundo [...], participa na evolução e da revolução do homem em

³⁸ POPPER, K. R. **O racionalismo crítico na política**. 2.ed. Brasília: UnB, 1994. 74p.

direção à sua história. Como elemento organizador, [...] que referencia o homem ao seu destino”, afirma Barreto (1994, p.1).

Buckland (1991) apresenta a informação como coisa, o que possibilita que se possam criar ou estabelecer regras quanto a inferências sobre a informação estocada. Retoma-se, portanto a noção de informação como um elemento mensurável, tratável, um bem.

Robredo (2003, p.9) relaciona algumas características da informação, com uma ênfase na abordagem como “coisa” remetendo a um ciclo informacional. Essas características apresentadas a seguir parecem ser aquelas que a CI pode levar em consideração:

A informação é suscetível de ser: registrada (codificada) de diversas formas, duplicada e reproduzida *ad infinitum*, transmitida por diversos meios, conservada e armazenada em suportes diversos, medida e quantificada, adicionada a outras informações, organizada, processada e reorganizada segundo diversos critérios, recuperada quando necessário segundo regras preestabelecidas.

Vale salientar que a simples disponibilidade da informação não garante sua socialização, é necessária sua aceitação pelo indivíduo, como bem afirma Barreto (1994) e Francelin (2003) quando se referem ao receptor.

A importância da informação para a sociedade está relacionada na atualidade à chamada sociedade da informação. Trata-se de uma sociedade, em que a comunicação da informação é fundamental para a tomada de decisão em quaisquer setores, “incluindo a superestrutura política, os governos [...], a cultura e as artes, a ciência e a tecnologia, a educação [...], a saúde, a indústria, as finanças, o comércio e a agricultura, a proteção do meio ambiente [...], etc.”, e em que “A sociedade passa progressivamente a funcionar em rede” (MIRANDA, 2000, p.80). Como afirma Barreto (1996, p.2), a informação “deixa de ser, unicamente, medida de organização por redução de incerteza, passa a ser a própria organização em si”, ou seja, deixa de ser um dos elementos de funcionamento das organizações e passa, em muitos casos, a ser o elemento mais importante.

Essa sociedade da informação torna-se possível através do uso das TICs, e pode levar ao que Lévy chama de o espaço do saber. Neste espaço, as TICs colocariam os problemas dos vínculos sociais sobre novas bases e abririam possibilidades não somente para pensarmos coletivamente a aventura humana, mas, principalmente, para influenciá-la através de novas formas de pensamento e relacionamento que ajudam a emergir na humanidade as inteligências coletivas (LÉVY, 1999).

Ao se pensar na utopia de uma Sociedade da Informação realmente inclusiva, vale resgatar o ideal de Henri de La Fontaine e Paul Otlet, que ao criarem a União das Associações Internacionais de Bibliografia, em 1910 (MATTELART, 2006), decidiram no primeiro congresso mundial das associações criarem a revista “*La Vie Internationale*”. Afirmaram que o aspecto social dessa organização era “contribuir para desenvolver as relações além das fronteiras, para crescer a solidariedade humana e para assegurar a paz entre as nações” (LA FONTAINE; OTLET, 1912, p. 233³⁹ citado por FREIRE, 2005).

A sociedade da informação ao fazer uso das tecnologias, principalmente a Internet e a Web, deveria estar diminuindo as desigualdades, embora nem sempre isso aconteça. A disponibilidade dos artefatos tecnológicos não é suficiente e a apropriação de informação leva também em conta aspectos econômicos (condições de possuir os equipamentos) e cognitivos (condições de transformar a informação em conhecimento). Isso é retratado à luz da pirâmide das necessidades humanas proposta por Maslow (1970)⁴⁰ citado por Barreto (1994), em que Barreto (1994) afirma que a demanda de informação é maior, quando as necessidades mais fundamentais estão satisfeitas.

Existe, portanto, o desafio de tornar efetiva a “sociedade da informação”, para que o ideal de Otlet e La Fontaine seja alcançado. A Internet e a Web podem colaborar neste empreendimento de forma significativa.

2.5 Natureza interdisciplinar

O contexto científico pós-moderno trouxe consigo a diminuição da fragmentação dos saberes, isto é, o pensamento temático se sobressai em detrimento do pensamento disciplinar. Nesse sentido, o modelo disciplinar postulado pela ciência moderna já não responde aos problemas sociais, exigindo outras abordagens. Surgem, portanto, termos como interdisciplinaridade, multidisciplinaridade, pluridisciplinaridade, transdisciplinaridade, além dos *Morinrianos*⁴¹, extradisciplinaridade, polidisciplinaridade, ecodisciplinaridade, metadisciplinaridade.

Para o presente trabalho, o termo mais relevante é interdisciplinaridade, que segundo a literatura é

³⁹ LA FONTAINE, H.; OTLET, P. La vie internationale et l’effort pour son organization. **La Vie Internationale**, v.1, n.1, 1912.

⁴⁰ MASLOW, A. H. **Motivation and a personality**. New York: Harper, 1970. Abraham Harold Maslow, psicólogo americano.

⁴¹ Neologismo referente a seguidores das idéias de Edgar Morin.

uma característica natural da CI, pois “incorpora os resultados de várias disciplinas, tomando-lhes de empréstimo esquemas conceituais de análise a fim de fazê-los integrar, depois de havê-los comparado e julgado” (JAPIASSU, 1976, p.32). Japiassu (1976, p.32) aponta algumas justificativas para o interdisciplinar, a saber:

1) proporciona trocas generalizadas de informação e de críticas, contribuindo, dessa forma, para uma reorganização do meio científico [...]; 2) amplia a formação geral de todos quantos se engajam na pesquisa científica especializada [...]; 3) questiona a possível acomodação dos cientistas em seus pressupostos implícitos [...]; 4) prepara melhor os indivíduos para a formação profissional [...]; 5) prepara e engaja os especialistas na pesquisa em equipe, fornecendo-lhes os instrumentos conceituais para que saibam analisar as situações e colocar os problemas [...]; 6) assegura e desenvolve a educação permanente que permite aos pesquisadores o prolongamento constante [...]

O fenômeno interdisciplinar segundo Japiassu (1976, p.42) “tem dupla origem: uma interna, tendo por característica essencial o replanejamento geral do sistema das ciências, que acompanha seu progresso e sua organização; outra externa, caracterizando-se pela mobilização cada vez mais extensa dos saberes convergindo em vista da ação”. Isso evidencia uma característica da crise do modelo científico chamado moderno e das características das ciências chamadas sociais, demonstrando uma relação clara com a CI.

Interdisciplinaridade pode ser definida então como o agrupamento de disciplinas com nível hierárquico e objetivos múltiplos, ou seja, cooperação de disciplinas, onde há um enriquecimento mútuo, com uma coordenação (JAPIASSU, 1976; ORRICO, 1999). Trata-se de uma característica inerente à forma pós-moderna de fazer ciência, se pautando para a ação. Com a interdisciplinaridade surgem alguns outros termos, a saber:

- Disciplinaridade: “[...] conjunto sistemático e organizado de conhecimentos que apresentam características próprias nos planos do ensino, de formação, dos métodos e das matérias”, em um domínio homogêneo de estudo (JANTSCH, 1972)⁴² citado por (JAPIASSU, 1976, p.72).
- Multidisciplinaridade: grupo de disciplinas com objetivos múltiplos, sem cooperação e sem coordenação, isto é, sem interferência de umas sobre as outras (JANTSCH, 1972) citado por

⁴² JANTSCH, E. Vers J’interdisciplinarité et la transdisciplinarité dans l’enseignement et J’innovation. In: OCDE. **L’interdisciplinarité**. 1972. p.108-109.

(JAPIASSU, 1976). Acontece na formação de grupos para a solução de problemas que necessitam de conhecimentos de diversas disciplinas.

- Pluridisciplinaridade: grupo de disciplinas com objetivos múltiplos, com cooperação, onde há enriquecimento das disciplinas envolvidas, mas não há coordenação. (JANTSCH, 1972 citado por JAPIASSU, 1976).
- Transdisciplinaridade: para Nicolescu (2000, p.15)⁴³ citado por Bicalho e Borges (2003), ela está relacionada “àquilo que está ao mesmo tempo entre as disciplinas, através das diferentes disciplinas e além de qualquer disciplina. Seu objetivo é a compreensão do mundo presente”.

González de Gómez et al. (2003, p.4) apresentam a seguinte definição de transdisciplinaridade:

[...] junção de um programa de pesquisa e de um programa de ação, em torno de questões ou problemas contextualizados, onde os processos de aprendizagem e descoberta são articulados numa matriz intersubjetiva plural e diversificada, formada por uma rede de sujeitos individuais e coletivos que possuem objetivos comuns e alguma forma de aliança ou parceria; [...] compartilhamento da produção de conhecimentos, as quais transgridem fronteiras e estruturas disciplinares, organizacionais e setoriais [...]

A CI, como uma ciência social, deve objetivar as soluções dos problemas informacionais, uma vez que é evidente a influência da informação no desenvolvimento da sociedade, sendo assim, ela deve buscar contribuições de áreas variadas, ou mesmo procurar interagir, ou seja, não apenas utilizar o conhecimento, mas também trabalhar em conjunto. Quanto às relações interdisciplinares da CI, além das já apresentadas com a Biblioteconomia e a Documentação, que colaboram inclusive para a origem da área, existem outras relações, propostas por autores diversos e apresentados nos Quadros 6 e 7 abaixo.

⁴³ NICOLESCU, B. Um novo tipo de conhecimento – transdisciplinaridade. In: NICOLESCU, B. et al. **Educação e transdisciplinaridade**. Brasília: UNESCO, 2000. p.13-29. (UNESCO)

Quadro 6. Relações interdisciplinares da CI na proposta de Lena Vânia Ribeiro Pinheiro

Ciência da Informação	ATIVIDADES	DISCIPLINAS
	Estudos de Usuários	Psicologia, Antropologia, Sociologia, Comunicação
	Sistema de Recuperação da Informação	Arquivologia, Biblioteconomia, Informática , Museologia
	Representação da Informação	Arquivologia, Biblioteconomia, Museologia, Filosofia, Linguística
	Biblioteconomia	Estatística, Matemática, História da Ciência
	Comunicação científica e tecnológica	Sociologia da Ciência, História da Ciência, Comunicação
	Automação	Arquivologia, Biblioteconomia, Informática , Museologia
	Administração de Sistemas de Informação	Biblioteconomia, Administração, Museologia, Estatística, Arquivologia, Informática
	Informação, Cultura e Sociedade	Biblioteconomia, História, Museologia, Educação, Arquivologia, Antropologia, Sociologia
	Epistemologia da CI	Filosofia, Epistemologia, Filosofia da Ciência
	Divulgação Científica	História da Ciência, Jornalismo científico, Comunicação
	Economia da Informação	Economia, Estatística, Administração
	Redes e Sistemas de Informação	Arquivologia, Biblioteconomia, Informática , Museologia

Fonte: Adaptado de Pinheiro (1999).

A partir do Quadro 6, podemos notar que a CI realiza algumas de suas atividades (segunda coluna) utilizando o conhecimento das diversas disciplinas citadas (terceira coluna). É o que foi denominado de cooperação com coordenação. As disciplinas dão sua cooperação tendo em vista a necessidade da disciplina coordenadora (CI), assim temos a interdisciplinaridade. Através dos grifos nas colunas do Quadro 6, pretende-se demonstrar possíveis aproximações da CI com a WS. Vale salientar que, de uma forma ampla, todas as atividades (segunda coluna do Quadro 6) da CI são necessárias e possíveis no ambiente Web de forma geral, foram destacadas apenas aquelas que entendemos serem fundamentais para o desenvolvimento da WS, na perspectiva já salientada: recuperação e representação da informação, pois não podemos negar que há estudos e possibilidades de pesquisas epistemológicas, estudos de usuários, economia da informação no ambiente Web, conseqüentemente, e no futuro na WS também.

Já, a proposta interdisciplinar da CI apresentada no Quadro 7 abaixo é defendida por Saracevic (1996).

Quadro 7. Relações Interdisciplinares da CI na proposta de Saracevic

Ciência da Informação	DISCIPLINAS
	Biblioteconomia
	Ciência da Computação
	Ciência Cognitiva (Inteligência Artificial)
	Comunicação

Fonte: Baseado em Saracevic (1996).

Fazendo uma comparação com as propostas apresentadas, podemos dizer que Saracevic (1996) estabelece uma relação mais simples, colocando a Biblioteconomia como uma disciplina, como já mostrado, além de vir com um elemento novo, a “Ciência Cognitiva”, especificamente a Inteligência Artificial, já que a Ciência Cognitiva ou “Ciências Cognitivas” é um conjunto de disciplinas que se estabelece a partir do estudo da mente humana. Nesse sentido, podemos incluir a Filosofia, a Lingüística e outras disciplinas citadas no Quadro 6, que, além disso, indica a Biblioteconomia como uma disciplina e também como uma atividade da CI.

Assim, a partir das propostas interdisciplinares apresentadas nos Quadros 6 e 7, propõem-se, no Quadro 8, as relações interdisciplinares da CI que embasam a presente pesquisa e que justificam a relação da CI com a WS, no que diz respeito à representação da informação e do conhecimento, bem como a RI. Vale salientar que não estão inclusas a Biblioteconomia e a Documentação, uma vez que entende-se atualmente que, essas são áreas de atuação profissional e não disciplinas independentes, apesar de colaborarem com a origem da CI.

Quadro 8. Relações interdisciplinares propostas na presente pesquisa

CI e WS	DISCIPLINAS
	Ciência da Computação
	Ciência Cognitiva – Inteligência Artificial
	Lingüística – Terminologia

Sendo assim, segue a partir de agora uma apresentação de cada disciplina citada no Quadro 8, com base na contribuição que à luz da CI elas podem oferecer no desenvolvimento da WS.

2.5.1 Ciência da Computação

Segundo Saracevic (1996, p. 50), “a base da relação entre CI e CC reside na aplicação dos computadores e da computação na recuperação da informação, assim como nos produtos, serviços e redes associados”. Podemos defini-la como a ciência que realiza o “estudo sistemático dos processos algorítmicos que descrevem e transferem informações: sua teoria, análise, desenho, eficiência, implementação e aplicação” (DENNING et al., 1989, p.12), através de uma relação entre ciência, engenharia e matemática (DENNING, 2005, p.27).

Para um melhor entendimento da CC, Denning et al. (1989) indicam nove subáreas dessa disciplina, a saber:

- Algoritmos e estruturas de dados: nessa subárea define-se o que a máquina pode e não pode fazer, isto é, a forma como o computador vai trabalhar;
- Linguagens de programação: trata-se de anotações para que as máquinas executem algoritmos, ou seja, é o estabelecimento de códigos entre o que um programador deseja que a máquina efetue com os algoritmos e estruturas de dados próprios da máquina;
- Arquitetura: cuida de métodos que organiza o hardware e o software de forma confiável para um funcionamento eficiente do sistema;
- Computação numérica e simbólica: preocupa-se com métodos gerais para resolução de equações, que são modelos matemáticos responsáveis pela precisão e eficácia de um certo sistema computacional;
- Sistemas operacionais: são os mecanismos de controle que permitem coordenar os recursos na execução dos programas;
- Metodologia e engenharia de softwares: trata de programas e sistemas de software, respondendo a perguntas como: como se prova que um programa ou sistema conhece suas especificações;

como se desenvolvem especificações que não omitem casos importantes e podem ser analisadas para segurança; como sistemas de software evoluem por gerações diferentes.

- Base de dados e sistemas de recuperação da informação: trata da organização de grandes bases de dados, respondendo a questões como: que modelos deveriam ser usados para representar elementos de dados e suas relações; como estas transações podem interagir efetivamente com o usuário; que arquiteturas de máquina conduzem a recuperação eficiente e atualização.
- Inteligência Artificial e Robótica: busca modelar o comportamento humano e animal e responder às seguintes perguntas: o que são modelos básicos de comportamento e como construímos máquinas que os simulam; até que ponto a inteligência é descrita por avaliação de regra, conclusão, dedução, e padrão de computação.
- Interação homem-computador: cuida da transferência adequada de informação entre humanos e máquinas, com estruturas que reflitam concepções humanas, respondendo a perguntas como: quais métodos são efetivos para contribuir na recepção ou produção; como o erro humano subsequente pode ser minimizado em uma certa operação?

Parece haver conflitos entre a CI e essas subáreas da CC. Por exemplo, no caso da subárea “Bases de dados e sistemas de recuperação da informação”, a CC tem seu foco na estruturação e organização do programa, ou seja, ela elabora a infra-estrutura para que isso aconteça, não levando em consideração diretamente um documento ou uma certa fonte de informação, partindo da escolha do algoritmo e estrutura de dados adequada, usando uma linguagem de programação pertinente. Na CI quando pensa-se em bases de dados e recuperação da informação, parte-se da necessidade de uma comunidade específica e do tipo de material que formará essa base para ser recuperado. Busca-se, em seguida, a infra-estrutura tecnológica adequada. Saracevic (1996, p.50) afirma que “[...] a ciência da computação trata de algoritmos que transformam informações enquanto a CI trata da natureza mesma da informação e sua comunicação para uso pelos humanos. Ambos os objetos são inter-relacionados e não competidores, mas complementares”.

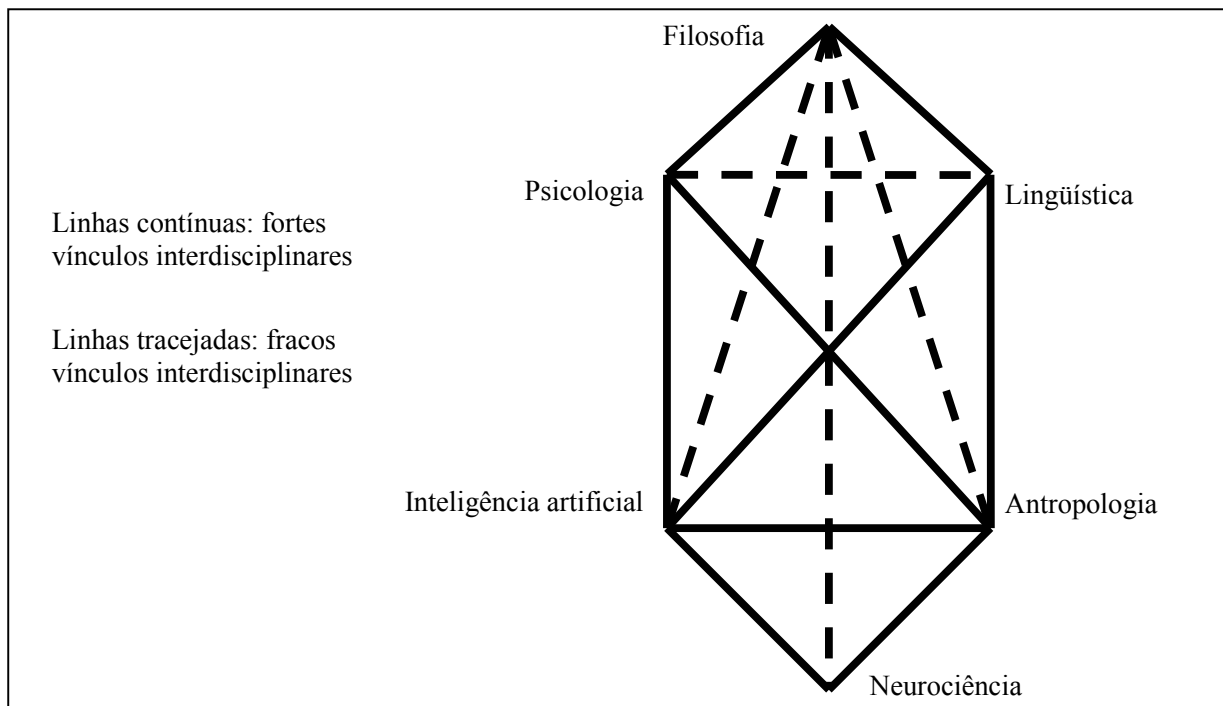
Essa complementação é inevitável, principalmente no ambiente Web. A aproximação dessas disciplinas ajudará no desenvolvimento da WS no que diz respeito à representação e recuperação da informação.

2.5.2 Ciências Cognitivas: Inteligência Artificial

A cognição é o conjunto de processos de “reflexão das leis e das propriedades de objetos e fenômenos da realidade objetiva na consciência humana” (FOGL, 1979⁴⁴ citado por BRASCHER; CAFÉ, 2008). Segundo Hjørland (c2005), a “Ciência Cognitiva” ou as “Ciências Cognitivas” são um campo de pesquisa interdisciplinar com raízes na psicologia cognitiva, lingüística, inteligência artificial e filosofia. Casti (1989)⁴⁵ citado por Saracevic (1996, p.51) acresce às disciplinas citadas: a Antropologia, a Neurofisiologia e a CC, afirmando que a união dessas áreas busca através do uso computador entender os segredos da mente.

A Figura 3 ilustra as interfaces da Ciência Cognitiva, segundo Gardner (1996, p.52)⁴⁶ citado por Lima (2003, p.81).

Figura 3. Hexágono cognitivo e seus seis campos constituintes.



Fonte: Gardner (1996, p.52)⁴⁷ citado por Lima (2003, p.81).

⁴⁴ FOGL, J. Relations of the concepts 'information' and 'knowledge'. **International Fórum on Information and Documentation**, The Hague, v.4, n.1, p. 21-24, 1979.

⁴⁵ CASTI, J. L. **Paradigms lost**: images of man in the mirror of science. New York: William Morrow, 1989.

⁴⁶ GARDNER, H. **A nova ciência da mente**: uma história da revolução cognitiva. São Paulo: EDUSP, 1996.

⁴⁷ Idem 46.

A partir da figura apresentada, parece correto afirmar que a filosofia e a lingüística lidam com a representação da informação, a antropologia e a psicologia buscam entender os indivíduos em relação ao seu meio (LOUREIRO, 1999), a neurociência estuda o sistema de cognição humana, e a Inteligência Artificial (IA) busca reproduzir e entender como o ser humano pensa.

Segundo Lima (2003, p.81), a Ciência Cognitiva se fundamenta a partir de três pontos:

Desenvolvimento da psicologia do processamento da informação, na qual a meta era especificar o processamento interno envolvido na percepção, linguagem, memória e pensamento; a invenção dos computadores e as tentativas de projetar programas que pudessem fazer tarefas que as pessoas fazem; desenvolvimento da teoria da gramática generativa e outras derivações da lingüística.

O interesse desse trabalho em relação às disciplinas que formam a Ciência Cognitiva se prende à IA.

Inteligência Artificial

Alguns autores entendem que a IA é um braço da CC (DENNING et al., 1989; HJORLAND, c2005). Já Saracevic (1996), apresenta a IA a partir da perspectiva da Ciência Cognitiva. Teixeira (1990) afirma que somente com o surgimento dos computadores modernos foi possível um projeto efetivo para a IA. Sua idéia principal é a criação de “máquinas pensantes”, ou seja, fazer com que os “mecanismos” da mente humana se reproduzam a partir da elaboração de programas de computador (DENNING et al., 1989; TEIXEIRA, 1990), por isso aqui ela é apresentada sob a ótica da Cognição, pois a base para a geração dessas “máquinas pensantes” é a mente humana e seu funcionamento.

A partir da assertiva anterior, podemos afirmar que a IA colabora para que a CC produza programas que apresentem cada vez mais familiaridade com a forma de pensar do homem.

Teixeira (1990, p.24), a respeito do surgimento da IA, comenta que:

Os resultados do Simpósio de Hixon (1948) não teriam sido tão surpreendentes se não levassem, através de uma intuição verdadeiramente criadora, a se estabelecer uma analogia entre o cérebro humano e os computadores. Essa analogia certamente foi o produto do encontro entre psicólogos, neurofisiólogos e engenheiros

eletrônicos que perceberam que o modo como estão dispostas as células no nosso cérebro (neurônios), ligadas através de fios nervosos minúsculos, é semelhante ao circuito elétrico de um computador eletrônico. Estava aberto o caminho para se dizer que a mente humana pode ser imitada por um computador. Com isto nascia a nova disciplina, a Inteligência Artificial.

A IA pode trazer muitas contribuições para a CI em termos de organização da informação e do conhecimento, pois sua influência na construção dos programas de computador, responsáveis pelas inferências em sistemas de informação como a Web são claros. Isso pode ser exemplificado quando ao se fazer uma pesquisa em um buscador web, digitamos o termo com algum erro, e o sistema nos apresenta o possível termo correto. Certo é que, essa relação deve ser estreitada, e a WS pode proporcionar isso. Isso foi mostrado por Bush (1945), quando da idealização do MEMEX, que é a informatização da forma humana de pensar e associar documentos e informações.

Evidente também é que a IA como pensada no seu princípio não se tornou realidade. Sua relação com a CC tem possibilitado facilidades para os usuários da web, mas ela acontece no nível sintático, ou seja, a IA ainda não possibilitou às máquinas entenderem e levarem em consideração o contexto; as decisões ainda cabem aos indivíduos.

2.5.3 Lingüística: Terminologia

A Lingüística pode ser definida como o estudo científico da linguagem, mas Dubois et al. (1999, p.394) afirmam que ela não se limita apenas a essa atividade, acrescentando também como aspectos de seu interesse de estudo: “a fala, o discurso, as relações da língua com o indivíduo e o mundo”. Segundo Stumpf e Weber (2003), a relação dessa disciplina com a CI se dá pela necessidade da representação da informação recorrer a processos verbais, na busca do controle de termos para permitir o armazenamento e a recuperação adequados da informação. Mendonça (2000) reforça essa afirmação, dizendo que a relação dessas áreas é estreita e se dá pela necessidade da CI de realizar as análises documentárias, utilizando métodos e processos lingüísticos para descrever o conteúdo dos documentos.

Sendo assim, parece propícia uma relação da Lingüística com a CI, pelas indicações já citadas e também pela afirmação de Smit (1986) sobre a relação da CI com a linguagem. Reforça ainda a

relação posposta, o tema desse estudo: a WS, que é qualificada como uma web inteligente, através da semântica. A semântica como área da Lingüística é “um meio de representação do sentido dos enunciados. A teoria semântica deve explicar as regras gerais que condicionam a interpretação semântica dos enunciados” (DUBOIS et al., 1999, p.527).

O estudo de Mendonça (2000), deixa a relação dessas áreas evidente, a partir da análise de artigos de alguns periódicos científicos da área da CI, afirmando:

Observou-se que os artigos nesta interface lingüística e ciência da informação estavam inseridos na grande área da **organização do conhecimento e representação da informação**, caracterizados pelos: fundamentos teóricos; sistemas de classificação e tesauros, estrutura e construção; classificação e indexação, métodos; representação do conhecimento por linguagem e terminologia; classificação e indexação aplicadas; sistemas de classificação universal e tesauros; ambiente da organização conhecimento. Dentro deste contexto, identificaram-se sete grupos principais, que são: • teórico – abordagem textual; • quantitativo – lingüístico e bibliométrico; • **temático – processamento intelectual, abordagem semântica, conceitual e terminológica**; • **aplicativo – projetos e modelos de indexação automática e linguagem natural**; • ensino – relações curriculares; • **tecnológico – sistemas especialistas e inteligência artificial**; • **normativo – lingüística e classificação decimal universal** (MENDONÇA, 2000, p.52, grifo nosso).

A partir dos destaques acima, podemos considerar que essa relação se concentra na organização do conhecimento e representação da informação (primeiro grifo), alicerce para uma eficaz recuperação da informação, além de prestar-se aos estudos teóricos da CI, organização dos eixos temáticos do conhecimento e da informação, a partir da representação, através de aplicações de indexação automática e linguagem natural (uma relação que ocorre entre a web, que necessita de indexação automática, e o seu usuário, que tem como referência a linguagem natural), no desenvolvimento tecnológico (fundamental para a web, pois sistemas especialistas e inteligência artificial serão necessários para seu desenvolvimento no caminho à WS) e em atos normativos, necessários no que diz respeito à organização da informação, padronização e normalização (segundo grifo).

Mostrada a relação indiscutível entre as áreas, sabe-se que a CI possui, segundo Tálamo e Lara (2006, p.204), um subdomínio denominado Lingüística Documentária, onde “são estudadas as características da linguagem dos ambientes informacionais que combinam as referências da produção informacional, os objetivos institucionais e os elementos cognitivos e comunicacionais de grupos de usuários”. Esse subdomínio se desenvolve principalmente a partir da apropriação dos

conceitos da Lingüística Estrutural, da Semiótica, da Terminologia e da Lógica Formal. Mas segundo Tálamo e Lara (2000, p.209), a Lingüística e a Semiótica não resolvem os problemas na delimitação de subconjuntos lingüísticos temáticos e funcionais necessários à Lingüística Documentária, ou seja, “não oferecem instrumental concreto à CI para a elaboração de suas linguagens, porque seus pontos de partida são especulativos. A Terminologia, porém, permite dar conta desses elementos nos seus aspectos teórico-metodológicos e materiais”.

A Terminologia fornece referências teórico-metodológicas necessárias à compreensão e formulação das redes lógico-semânticas, observando-se uma relação estreita entre a norma documentária para elaboração de tesouros (ISO 2788) e as normas terminológicas (ISO 704 e ISO 1087). As terminologias concretas, por sua vez, fornecem o referencial interpretativo para os descritores (os termos funcionam como operadores de sentido) e para a linguagem documentária de um modo geral: oferecem procedimentos para a delimitação dos universos-foco, orientam a identificação dos domínios e dão sustentação à construção das redes relacionais, seja pela definição dos termos, seja pela identificação das características para compor as estruturas de relacionamento (TÁLAMO, LARA, 2000, p.210).

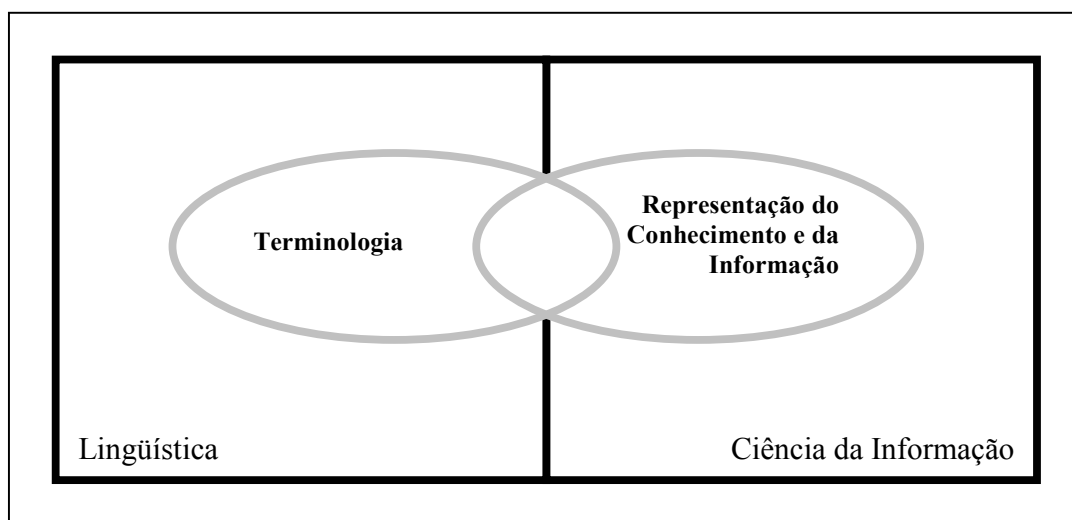
Cabré (1995) apresenta três concepções da Terminologia que podem contribuir para um melhor entendimento da área e da aplicação feita pela CI para a organização da informação. São elas: a) a disciplina; b) a prática; c) e o produto gerado pela prática. Como disciplina, a Terminologia se ocupa dos termos especializados, como prática cuida do conjunto de diretrizes e princípios que orientam a compilação dos termos e como produto da prática tem-se o conjunto dos termos de um domínio específico.

Como aplicação, a Terminologia possui duas finalidades: a representação e a transferência. Na primeira ela serve a três disciplinas ou atividades: a documentação; a engenharia lingüística e a lingüística computacional; e as especialidades científico-tecnológicas (CABRÉ, 1995). Justifica-se, portanto, sua relação com o nosso trabalho, com uma ênfase na documentação, que se vale da terminologia com o intuito de representar os documentos para recuperação e também com a lingüística computacional, necessária para a criação dos programas que fazem os computadores funcionarem, assim como permitem a interação entre máquinas e usuários. No que diz respeito ao ambiente digital, Cabré (2005) salienta que o futuro dessa subárea da Lingüística está na disponibilidade dos recursos digitais e de ferramentas de tratamento da informação que se dá pela ampliação dos portais virtuais do conhecimento. Assim, a Terminologia se prestará cada vez mais às atividades de “confeções de resumos, representação gráfica do conhecimento, recuperação de

documentos e constituição de sistemas de tratamento automático da linguagem”. Assim, parece correto afirmar que para a CI ela pode oferecer subsídios na construção da Web rumo à WS.

Sendo assim, a relação interdisciplinar proposta parece justificar-se, sendo apresentada na Figura 4, a seguir:

Figura 4. Intersecção entre a Lingüística e a Ciência da Informação



A partir de uma análise da figura, podemos afirmar que a subárea específica da CI denominada “Representação do Conhecimento e da Informação” emprega parte da área da Terminologia, principalmente no que diz respeito a seus aspectos práticos e aplicados para realizar suas atividades de representação.

3 INFORMAÇÃO E CONHECIMENTO: recuperação, organização e representação

Esse capítulo buscará desenvolver alguns conceitos da CI referentes à informação e ao conhecimento: recuperação, organização e representação. RI é um termo recorrente na CI, já a recuperação do conhecimento trata-se de um termo mais raro, principalmente pela distinção dos conceitos informação e conhecimento.

Para Brascher e Café (2008) a organização da informação (OI) trata de um processo que abarca a descrição física e de conteúdos da informação, entendendo, portanto, essa última como um objeto, uma coisa. O produto desse processo é a representação da informação, “conjunto de elementos descritivos que representam os atributos de um objeto documental específico” (BRASCHER; CAFÉ, 2008, p.5), assim os objetos da organização da informação e da RI são os registros de informação.

A relação da organização do conhecimento com a representação do conhecimento (RC), segue o mesmo sentido. Enquanto a primeira é um processo que objetiva a geração de modelos de mundo, que se presta a compor abstrações da realidade, a segunda é o seu produto que permite a materialização desse modelo abstrato, mas sem uma ligação direta com as informações (BRASCHER; CAFÉ, 2008).

Esse capítulo trata, portanto, da recuperação e da representação da informação, assim como da RC, seus conceitos, distinções, métodos e instrumentos.

3.1 Recuperação da informação

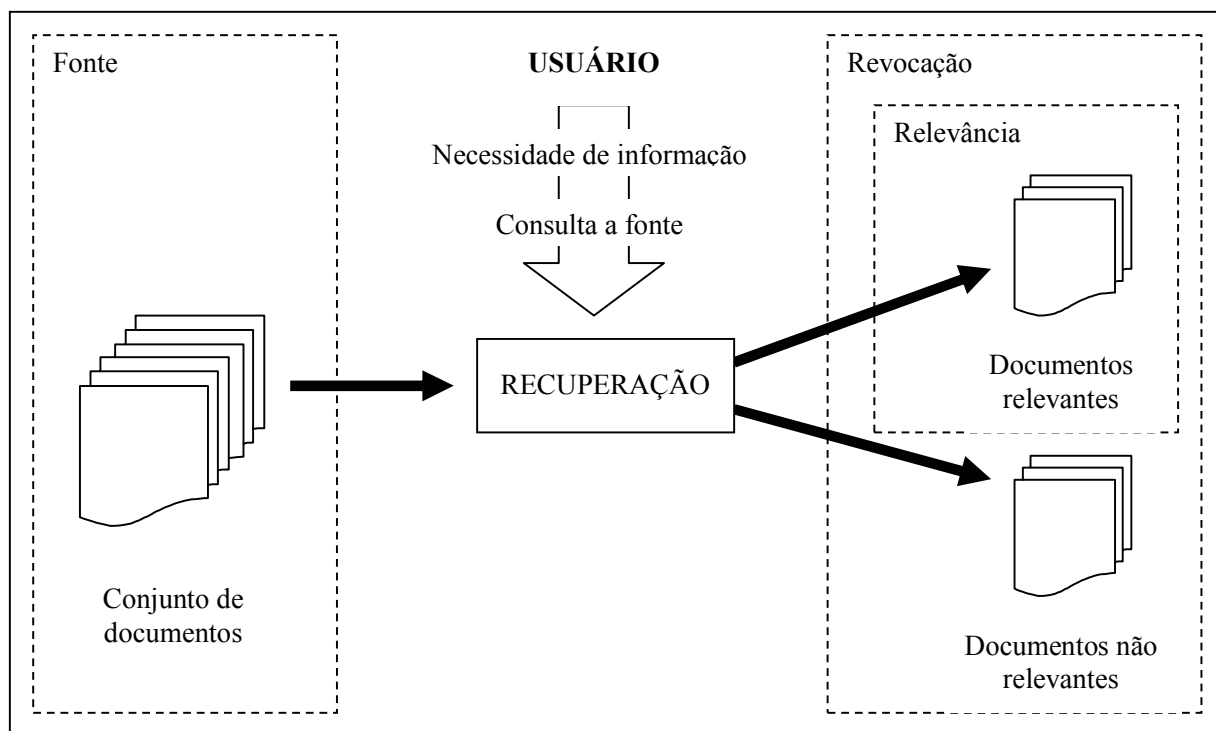
De forma breve, Meadow, Boyce e Kraft (2000) entendem que a RI consiste em encontrar uma informação desejada para alguém ou para si mesmo em alguma fonte. Nesse processo comunicacional está implícito, segundo esses autores, o conceito de seletividade, que é decorrência de investimento de esforço, tempo e dinheiro. Para o *“Online Dictionary for Library and*

Information Science”, a RI é descrita como o processo, os métodos e procedimentos utilizados para a revocação de registros de informação (REITZ, 2007).

Hjorland (c2005) afirma que Mooers foi o primeiro a utilizar o termo RI em 1951, definindo-o como o processo ou método pelo qual um usuário, necessitado de informação, pode converter essa necessidade em uma lista de documentos potencialmente úteis armazenados em alguma fonte, ou seja, é o processo de descoberta de uma informação específica armazenada. Diz respeito também aos aspectos intelectuais da representação da informação e sua especificação para procura, como também quaisquer sistemas, técnicas ou máquinas que são utilizados para efetuar a operação (MOOERS, 1951⁴⁸ citado por HJORLAND, 2005).

A Figura 5 abaixo ilustra o cenário da RI.

Figura 5. Cenário de recuperação da informação



Adaptado de Tolosa e Bordignon (2007, p.11).

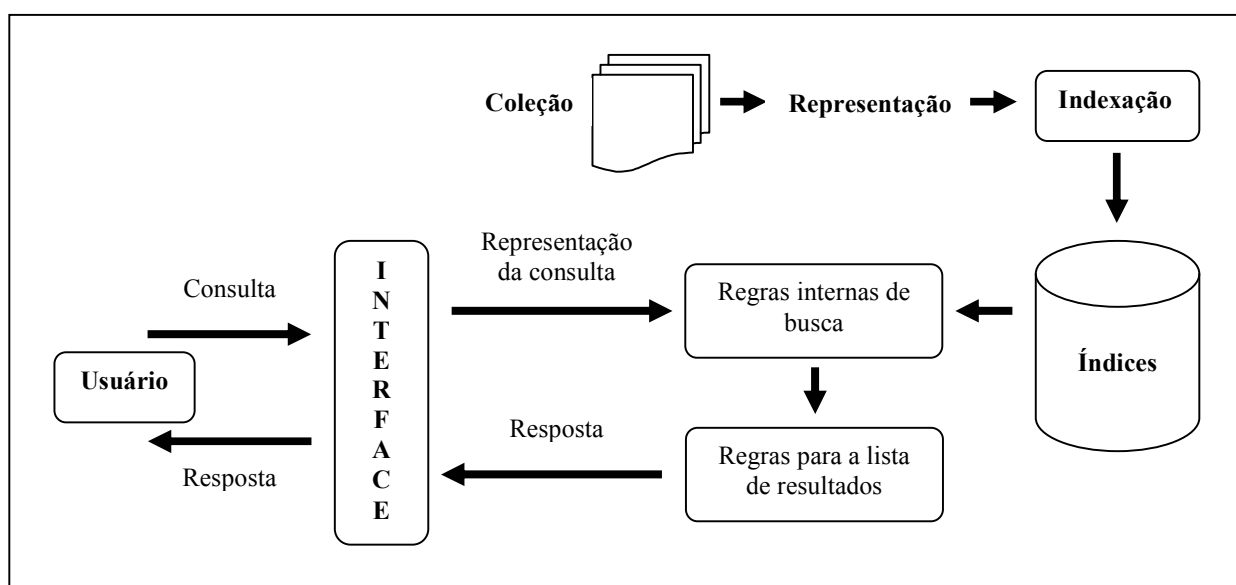
Meadow, Boyce e Kraft (2000) se referem à idéia intrínseca de seletividade presente na RI. A partir

⁴⁸ MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, Washington, n.2, p.20-32. 1951.

da Figura 10, pode-se perceber que diante de uma necessidade informacional, um certo usuário deve procurar a melhor maneira de elaborar a questão e escolher a fonte mais adequada que atenda à sua necessidade, assim como selecionar diante do resultado oferecido, as informações realmente relevantes.

Além da idéia de seletividade, outra que permeia a RI é a de sistema. Alguns autores como Foskett (1973), Tolosa e Bordignon (2007) se referem a “sistemas de recuperação da informação” (SRI), apontando a RI como uma parte desse sistema. Essa visão é válida, mas de forma geral, todos os processos de um SRI giram entorno da RI. A Figura 6 ilustra a arquitetura básica de um SRI.

Figura 6. Arquitetura básica de um sistema de recuperação da informação



Fonte: Tolosa e Bordignon (2007, p.12).

A RI pode ser estudada sob dois pontos de vista: o humano e o tecnológico. O primeiro diz respeito aos estudos do comportamento dos indivíduos na busca pela informação. O segundo refere-se a construção das estruturas que ajudarão na melhoria da qualidade da RI (TOLOSA e BORDIGNON, 2007).

Outra consideração importante a respeito da RI, é que esta não é necessariamente dependente dos computadores. Meadow, Boyce e Kraft (2000) fazem algumas considerações a esse respeito, descrevendo os métodos tradicionais de RI, como os catálogos e os índices manuais, instrumentos praticamente exclusivos das bibliotecas e outras unidades tradicionais de informação, como arquivos

e museus.

A RI é cercada por alguns conceitos que complementam seu entendimento, alguns deles já citados, como revocação e relevância. Também há outros, como precisão, redundância, exaustividade, etc., todos apresentados a seguir:

- **Revocação:** a proporção entre documentos pertinentes recuperados e o número total de documentos pertinentes no sistema sobre um determinado assunto (HJORLAND, 2005), ou seja, uma certa fonte de informação contém 100 registros sobre a história da CI como assunto, uma busca mostrará 50 registros, desses, 25 são pertinentes ao assunto, assim a revocação seria de 25%. Uma das grandes dificuldades da revocação, enquanto unidade de medida em RI é determinar o número total de documentos pertinentes em qualquer fonte de informação, mesmo as menores (REITZ, 2007).
- **Precisão:** enquanto a revocação é uma unidade de medida de RI referente ao total de documentos pertinentes na fonte de informação, a precisão diz respeito ao total dos documentos que foram recuperados. No exemplo indicado na definição acima, o índice de revocação foi de 25%, pois estava condicionado ao total de documentos pertinentes. A precisão nesse mesmo exemplo, é de 50%, pois ela diz respeito aos documentos considerados relevantes, dos 50 recuperados (REITZ, 2007).
- **Relevância:** se as informações recuperadas cumpriram sua finalidade de resolver a necessidade de um certo usuário elas são relevantes (FOSKETT, 1973; HJORLAND, 2005; REITZ, 2007). Responde, portanto, a seguinte pergunta: essa informação 'A', responde a questão 'B', do usuário 'C'? Isso não significa necessariamente, que essa relevância esteja ligada apenas ao conteúdo, deve-se levar em consideração outros aspectos, como: idioma (exemplo: um texto recuperado em idioma chinês para um usuário que nele não possua fluência, pode ser levado em consideração no que diz respeito à revocação e precisão, mas não será relevante para o usuário); valor ser pago (exemplo: um documento, dependendo do seu valor, pode não ser interessante para um certo usuário); etc.
- **Redundância:** um fator presente em qualquer SRI, que influencia diretamente a RI, seja na revocação, na precisão ou na relevância. O resultado de uma busca possui a característica da redundância quando apresenta informações idênticas repetidas.
- **Exaustividade:** é um conceito presente na RI, tanto na fase de entrada dos dados no SRI, como na de RI propriamente dita, pelo usuário. Por essa segunda vertente, o usuário faz uma busca em

todas as fontes possíveis, de todas as formas, com o objetivo de recuperar todas as informações pertinentes ao tópico que lhe interessa. Na primeira vertente, o catalogador ou indexador, busca identificar todos os assuntos presentes em um determinado documento (REITZ, 2007). Está diretamente ligada ao conceito de especificidade, descrito a seguir.

- Especificidade: “a extensão em que o sistema nos permite ser precisos ao especificarmos o assunto de um documento que estejamos processando. Quanto maior a especificidade, maior será a probabilidade de podermos alcançar alta relevância” (FOSKETT, 1973, p.12). Trata-se também de conceito presente tanto na entrada dos dados em um SRI, como no processo de RI por parte do usuário. Ou seja, mesmo que a necessidade de informação de um usuário seja bem específica, que a questão seja bem formulada junto ao SRI, como exaustividade na delimitação dos assuntos que cobrem a questão, a RI não será relevante, se não houver especificidade na representação da informação no sistema. Da mesma forma, se o usuário não for específico na sua questão, ele não terá grande relevância na RI, mesmo que a construção da representação documentária em certo sistema seja bem específica.
- Ruído: “determina a proporção de documentos irrelevantes encontrados no conjunto de documentos recuperados” (TOLOSA; BORDIGNON, 2007, p.41). Trata-se de um conceito importante, pois sugere questões como: por que o documento foi recuperado se ele não é relevante; a relevância, como um conceito praticamente exclusivo do usuário é, portanto, contextual, ou seja, para um outro usuário, mesmo com uma necessidade informacional parecida, o resultado do ruído pode ser diferente?
- Ambigüidade: esse conceito é proveniente da Lingüística, sendo a propriedade de certas frases apresentarem mais de um sentido ou interpretação (DUBOIS et al., 1999). Robredo (2003, p.236) sobre dois elementos intrínsecos à ambigüidade afirma o seguinte, “o significante, em termos simples, é a palavra, e o significado é a coisa ou idéia que ela representa”. A partir dessa idéia, Santos (2006, p.14), afirma que “a ambigüidade, portanto, pode ser entendida quando um único significante remete a mais de um significado”. Assim, pode-se afirmar que a ambigüidade parece ser um elemento relevante para a RI, se um termo tem significados diferentes para várias áreas do conhecimento, por exemplo, a busca por esse pode revocar muita informação e dificultar a identificação dos realmente relevantes para o usuário.
- Sinonímia: “dois termos são tidos sinônimos quando têm a possibilidade de se substituírem um ao outro num único enunciado isolado” (DUBOIS et al., 1999, p.555). Ou seja, é o contrário da

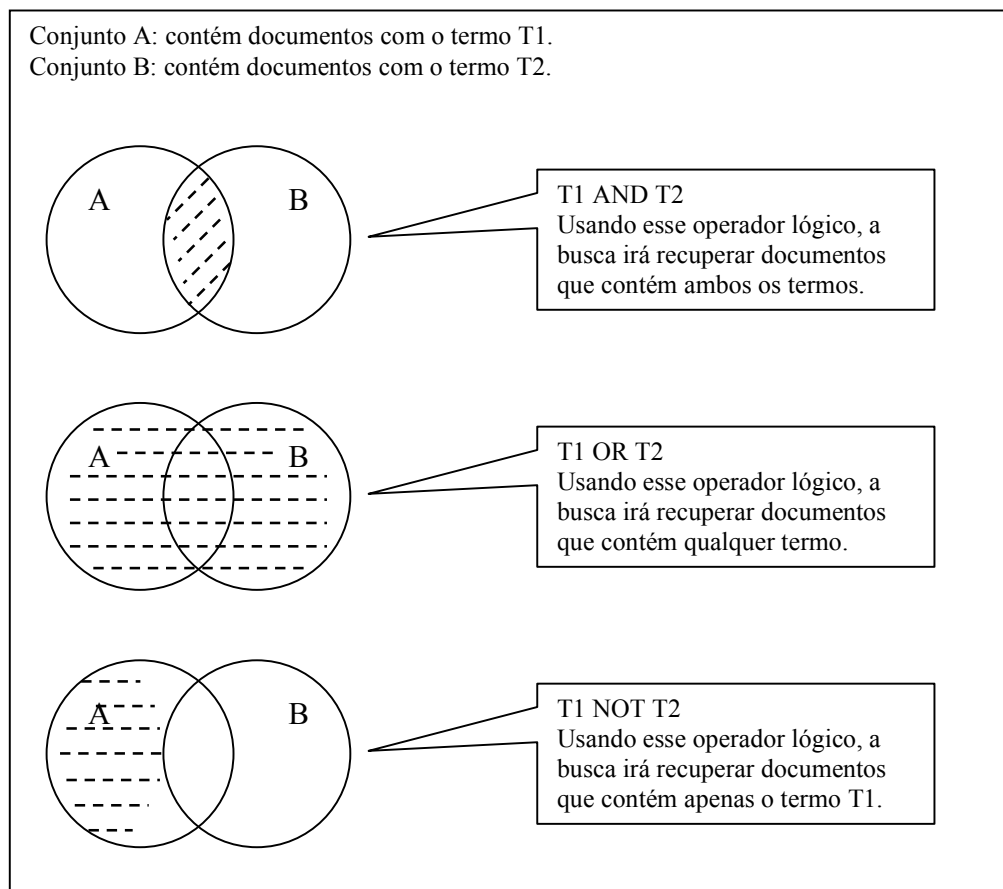
ambigüidade, acontece quando um significado tem mais de um significante. Para a RI isso é muito relevante. Quando da busca por parte do usuário, se ele quiser atingir a exaustividade, deve estar atento à sinonímia.

Existem alguns modelos clássicos que caracterizam a RI para documentos não estruturados, esses respondem a consultas através de buscas sobre uma estrutura de dados que representam o conteúdo/assunto dos documentos de uma coleção, por meio de listas de termos significativos (TOLOSA; BORDIGTON, 2007). Serão apresentados aqui dois deles, o booleano ou lógico e o vetorial:

- Booleano⁴⁹ ou lógico: a RI acontece através do uso dos operadores lógicos ou booleanos “AND”, “OR” e “NOT”, respectivamente “E”, “OU” e “NÃO” (ROBREDO, 2005; TOLOSA; BORDIGNON, 2007). Tais operadores ajudam o usuário a aumentar a relevância e precisão da busca pela informação desejada, além da diminuição do ruído. A Figura 7 exemplifica o uso dos operadores.

⁴⁹ De George Boole, irlandês que formulou a escrita algorítmica em 1854. Tal formulação possibilitou que a informação fosse construída como disciplina autônoma um século depois, permitindo o tratamento automático da informação (MATTELART, 2006).

Figura 7. Uso de operadores lógicos na recuperação da informação



Adaptação de Tolosa e Bordignon (2007).

- Vetorial: modelo de RI criado e desenvolvido por Gerard Salton aplicado a um SRI chamado SMART, nos anos de 1970 (MEADOW; BOYCE e KRAFT, 2000; TOLOSA; BORDIGNON, 2007).

Tolosa e Bordignon (2007, p.22) definem o modelo:

Conceitualmente, esse modelo utiliza uma matriz documento-termo que contém o vocabulário da coleção de referência e os documentos existentes. Na intersecção de um termo T e um documento D se armazena um valor numérico de importância do termo T no documento D; tal valor representa seu poder de discriminação. Assim, cada documento pode ser visto como um vetor que pertence a um espaço N dimensional, onde N é a quantidade de termos que compõem o vocabulário da coleção.

Existem ainda modelos para os documentos estruturados, esses complementam os modelos de consulta, utilizando a própria estrutura de cada documento para facilitação da recuperação. Esse modelo é aplicado principalmente para documentos digitais. Segundo Tolosa e Bordignon (2007), a estrutura dos documentos é feita através de marcação com etiquetas (tags), os principais padrões para a estruturação dessa forma são o SGML (Standard General Markup Language), o HTML, o PDF (Portable Document Format) e o XML (eXtensible Markup Language).

Assim, pode-se realizar a busca por diversas formas: termo no título, termo no corpo do texto, tipo de mídia, tipo de arquivo, período de publicação da informação, local de publicação, idioma, etc. Essas linguagens de marcação oferecem a possibilidade da disponibilização de documentos semi-estruturados na Web, o que teoricamente pode facilitar a especificidade de busca por parte do usuário.

Para tornar possível a RI, os processos de representação da informação, oriundos da Biblioteconomia e da Documentação, bem como o processo de representação do conhecimento, são imprescindíveis.

3.2 Representação da Informação e Representação do conhecimento

No presente trabalho, a informação é vista como coisa, elemento mensurável e administrável, conhecimento inscrito, registrado, autônomo como afirma Popper (1975), ou seja, gerado pelo conhecimento dos indivíduos, mas já “liberto” de uma ligação direta. Diz-se direta, pois certamente há uma ligação indireta entre informação e conhecimento, que são elementos distintos, mas certamente inseparáveis.

Essa distinção se estende à representação da informação e representação do conhecimento, visto que o termo informação e conhecimento têm sido usados como sinônimos, mesmo na área de CI. Esta situação pode ser observada por dois exemplos apontados por Brascher e Café (2008, p.7). No primeiro, as autoras citam Alvarenga (2006, p.5), apresentando suas reflexões entre colchetes.

Na representação primária, os produtos finais são constituídos de conceitos sobre os seres, formando o conhecimento, conceitos mais ou menos intensamente detalhados, codificados através de uma linha simbólica. Na representação

secundária, prática essencial nos sistemas informação documentais, esses mesmos conceitos constantes dos registros primários são sucintamente identificados em seus elementos constitutivos fundamentais, escolhendo-se os pontos de acesso fundamentais que garantem a representação desse conhecimento (documento) [*para nós informação*] para fins de futura recuperação. Neste caso, os conceitos constantes dos documentos, assim como suas superfícies de emergência, constituem-se em insumos para a representação secundária e devem ser identificados, requerendo-se que o profissional da informação, no processo de organização do conhecimento [*para nós organização da informação*], proceda à identificação dos elementos de descrição e temáticas que poderão vir a ser buscados pelos usuários potenciais do sistema de informação.

Apesar disso, uma afirmação de Alvarenga (2006) serve de base para a proposta de Brascher e Café (2008), ainda que denominando isso de RC, quando afirma que “as informações nesse tipo de representação compreendem compactações que tentam descrever as características do documento, refletindo sua origem e conteúdo, facilitando sua recuperação” (ALVARENGA, 2006, p.6).

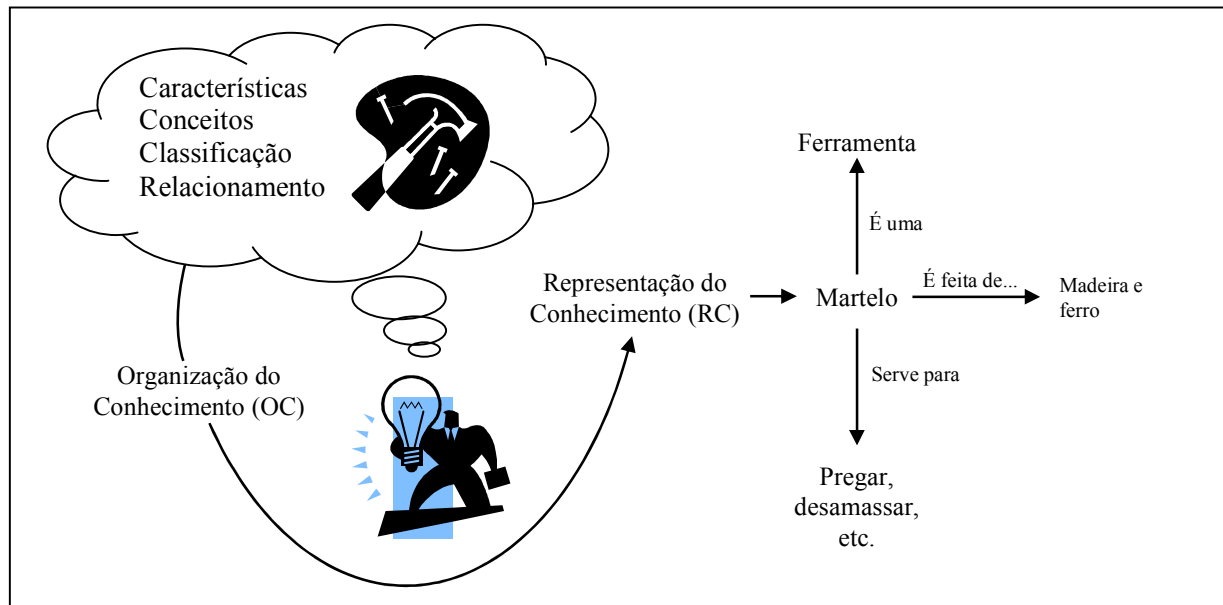
No segundo exemplo, Brascher e Café (2008) discordam de uma afirmação de Hjørland (2008), quando esse afirma que processos de OC, na visão da CI, compreendem elaborar resumos, catalogar, classificar, indexar, etc. Para as autoras tais atividades se aplicam a objetos físicos, ou como elas chamam, objetos informacionais, sendo assim, são atividades de representação da informação e não do conhecimento (BRASCHER; CAFÉ, 2008).

Assim, pode-se afirmar, que a representação da informação, como produto do processo OI, tem sua base no conhecimento registrado, na informação enquanto conhecimento objetivo, ou seja, estocada, armazenada.

A partir dessa premissa, seguem alguns instrumentos e atividades que dizem respeito à organização e representação da informação. Eles são essencialmente ligados aos registros, sendo desenvolvidos conforme a necessidade de fazer com que os objetos informacionais sejam recuperados. Afinal, como coisa, a informação está acumulada em suportes e necessita de métodos próprios para seu controle e acesso, seja pela representação física, seja pela representação de seu conteúdo.

A RC como produto do processo OC, presta-se à criação de instrumentos que auxiliem os usuários de informação. Mesmo sendo algo distinto da representação da informação, a RC só existe em função dela. Sua ligação é indireta com a informação enquanto objeto, mas direta com os usuários e profissionais da informação, auxiliando os usuários na RI e os profissionais na representação da informação.

A Figura 8 abaixo ilustra a OC e a RC.

Figura 8. Cenário de organização e representação do conhecimento

Adaptação de Brascher e Café (2008, p.7).

Algo que pode facilitar o entendimento do que seja a RC é a teoria do conhecimento objetivo de Popper (1975). Nessa teoria, o filósofo, como já citado anteriormente, propõe três mundos:

- Mundo 1: das coisas físicas, o mundo material.
- Mundo 2: do subjetivo, dos pensamentos, dos modelos mentais criados pelos seres humanos a partir da interpretação dos objetos do mundo 1.
- Mundo 3: do conhecimento objetivo, da materialização dos pensamentos elaborados pelo mundo 2. Essa terceira proposta de Popper justifica este trabalho.

Assim, pode-se afirmar que, instrumentos e atividades de RC, visam facilitar a representação da informação. Porém, possuem autonomia, auxiliando também no entendimento de uma determinada área do conhecimento. Por exemplo, um dicionário de Linguística, é um instrumento de RC, pois não possui ligação direta com nenhum pacote informacional, mas pode auxiliar na representação de um conjunto de documentos da respectiva área, como também tirar dúvidas de um simples leitor. É um material do mundo 3, pois está registrado, mas atua principalmente sobre o mundo 2.

Serão apresentados agora alguns instrumentos e processos da representação da informação e representação do conhecimento.

3.3 Classificação

Classificar é essencialmente colocar em ordem ou agrupar. Trata-se de uma atividade que de forma geral e mesmo sem intenção, um indivíduo acaba por realizar em algum momento. Maniez (1993) afirma que a classificação pode ser entendida de forma ampla como a ação de distribuir em classes qualquer conjunto de objetos, como também o produto resultante da operação de entender e classificar um sistema coerente e estruturado, como classificar os minerais, os répteis, etc. Robredo (2005, p.128) afirma que, originalmente, “a classificação foi descrita como o processo mental de designar e ordenar o universo do conhecimento, para determinar, com base em suas relações, o lugar preciso e exato das coisas, num esquema organizado”.

Para a CI, a classificação, por meio de uma linguagem, divide um domínio da realidade em uma série ordenada de classes e subclasses, se necessário, ou seja, representa o conteúdo de um documento de forma sintética (SLYPE, 1991). Maniez (1993, p. 23) define classificar como “[...] a operação intelectual na qual o documentalista atribui a uma obra um índice correspondente a uma classe de matérias, utilizando uma linguagem de classificação (= uma classificação documental)”. Pode-se afirmar, que se trata de uma atividade de representação da informação, pois ao menos, para a CI, está diretamente ligada aos estoques informacionais, pois se fala sempre em classificar documentos.

Outros termos básicos que permeiam a classificação segundo Maniez (1993) e que parecem interessantes para o presente trabalho são:

- Classe: conjunto de objetos que possuem características comuns. A subclasse é uma parte determinada da classe. Por exemplo, a classe dos advogados, é um conjunto de indivíduos que possuem curso de graduação em Direito, com licença de seu respectivo conselho profissional para exercer a função; trata-se também de classe com várias subclasses, como os advogados criminais, cíveis e trabalhistas.

- **Ordenação:** dispor objetos ou elementos de um conjunto seguindo uma ordem. Um exemplo ocorre nas bibliotecas, quando da colocação dos livros na estante, seguindo a ordem de classificação.
- **Notação:** código artificial composto símbolos numéricos e alfabéticos que representam o assunto da classificação.
- **Rubrica:** tradução em linguagem natural de um assunto correspondente a um índice na apresentação de uma classificação documental. Exemplo: 592 = invertebrados.
- **Tabelas de classificação:** documentos que apresentam de maneira ordenada e em forma de quadros uma série de índices e de rubricas de uma classificação documental.

Dos grandes sistemas de classificações desenvolvidos no âmbito da CI, principalmente na Biblioteconomia, citamos as seguintes, a partir de Maniez (1993) e Robredo (2005):

- **Classificação Decimal de Dewey (CDD):** sua organização é sistemática, contendo dez grandes classes, com base na divisão do conhecimento humano, possui estrutura decimal no sistema de notação. O Quadro 9 abaixo apresenta as classes principais da CDD e o Quadro 10, um exemplo de um conjunto de subclasses.

Quadro 9. Classes principais da Classificação Decimal de Dewey (CDD)

000 – Obras Gerais	500 – Ciências puras
100 – Filosofia	600 – Tecnologia
200 – Religião	700 – Artes
300 – Ciências Sociais	800 – Literatura
400 – Lingüística	900 – História

Fonte: Robredo (2005, p.130).

Quadro 10. Divisão da classe ‘Ciências Puras’ (500) da CDD

500 – Ciências Puras	550 – Geologia
510 – Matemática	560 – Paleontologia
520 – Astronomia e ciências afins	570 – Antropologia e biologia
530 – Física	580 – Botânica
540 – Química	590 – Zoologia

Fonte: Robredo (2005, p.130).

- Classificação Decimal Universal (CDU): linguagem de classificação derivada da CDD. Utiliza a idéia de classes principais e a notação decimal, mas apresentando subclasses mais avançadas, permitindo profundidade na análise dos assuntos e inovações nas subclasses e na notação. O Quadro 11 abaixo mostra as classes principais dessa linguagem de classificação e o Quadro 12 um exemplo de subclasses.

Quadro 11. Classes principais da Classificação Decimal Universal (CDU)

0 Generalidades, Documentação, Biblioteconomia	5 Matemática, Ciências Naturais
1 Filosofia, Ética, Psicologia	6 Ciências aplicadas, Medicina, Tecnologia
2 Religião, Teologia	7 Artes, Arquitetura, Esportes
3 Ciências Sociais, Direito Administrativo	8 Línguas, Literatura
4 Livre	9 Geografia, Biografia, História

Fonte: Robredo (2005, p.131).

Quadro 12. Divisão da classe ‘Ciência Aplicada. Medicina. Tecnologia’ (6) da CDU

6 CIÊNCIAS APLICADAS, MEDICINA, TECNOLOGIA		
61	Medicina. Ciências médicas	
62	Engenharia	
63	Agricultura. Silvicultura. Zootecnia. Caça. Pesca.	
	Agronomia. Ciências agrícolas. Tecnologia agrícola	
64	Ciência doméstica. Economia doméstica	
65	Administração e organização da indústria, do comércio e dos transportes	
66	Indústria química. Tecnologia química	
67/68	Indústrias e ofícios diversos. Tecnologia	mecânica
69	Construção	

Fonte: Robredo (2005, p.131).

Essas duas linguagens de classificação são as mais utilizadas na atualidade. Tratam-se de linguagens documentais de RC, que auxiliam na atividade de representação da informação. A classificação, portanto, é parte fundamental do tratamento da informação, precedendo, por vezes, a atividade de indexação, que será apresentada a seguir.

3.4 Indexação

A indexação é uma atividade de representação da informação, ou seja, ela está condicionada aos documentos. Mas, para que seja realizada, pode-se valer tanto de instrumentos de representação da informação como de instrumentos de RC. Robredo (2005) afirma que essa atividade dedica-se à identificação dos conceitos de um documento, ou seja, indica termos que representem o conteúdo total de um registro. Através das linguagens de indexação, Slype (1991) afirma que há a possibilidade da representação do conteúdo dos documentos de forma analítica, utilizando-se de oito a doze descritores de um certo tesauro, ou mesmo dezenas de palavras tomadas de seu título, resumo ou texto completo.

Por isso, é que talvez seja necessário que no ciclo documental, em um centro tradicional de informação, ou na Web, a classificação deva acontecer antes da indexação, pois a primeira relaciona um certo documento a um domínio de especialidade e a segunda, procura identificar todos os assuntos tratados nesse mesmo documento, servindo o domínio de baliza para a indexação.

A indexação na CI possui muitas formas de execução, bem como muitos tipos. Uma divisão considerável diz respeito à como a indexação é feita, ou seja, manual ou automática.

A indexação manual ou humana é um procedimento realizado por um indivíduo. A indexação automática é realizada por computador, por meio de procedimentos algorítmicos. O algoritmo trabalha em um banco de dados que contém representações de um certo documento em texto completo, registro bibliográfico ou partes de um texto (HJORLAND, 2008).

Slype (1991) salienta, porém, que a diferença entre essas duas formas de indexação, é que a manual, possibilita a leitura ou a consulta de uma linguagem de indexação, já a automática, é praticamente condicionada ao documento, ou seja, a partir de uma leitura automática de um documento, extraído os termos de partes importantes do texto, como título ou resumo, ou mesmo do texto completo. Parece correto afirmar, que a indexação humana em certa medida é mais semântica que a automática, porém, sujeita à interpretação do indexador, o que pode ser crucial para a qualidade do serviço.

Os principais tipos de linguagem de indexação, com base em Slype (1991), são:

- Linguagem natural: lista de palavras-chave (constituída por uma coleção não ordenada de palavras significativas, extraída de forma automática de um texto, eliminando as palavras não

significativas, como os artigos, conjunções, pronomes, preposições, numerais e certos verbos e advérbios); e lista de descritores livres (constituída por uma coleção não ordenada de conceitos selecionados por processo intelectual; apresentada em lista, em ordem alfabética, a partir de documentos de um sistema, conceitos que podem ser palavras ou expressões, ou seja, nesse caso, já um certo teor semântico no processo de indexação).

- Linguagem controlada: listas autorizadas de conceitos (representados por palavras ou expressões, estabelecidas anteriormente ao processo de indexação, ou seja, só os termos que constam na lista autorizada podem ser utilizados para indexar o documento e realizar consulta. Apresenta-se geralmente em ordem alfabética, sem controle hierárquico); e tesouros (lista estruturada de conceitos, geralmente de um domínio do conhecimento; instrumento de RC, utilizado para também como instrumento de indexação, haja vista, ser a OI, por vezes condicionada à OC).

Pode-se afirmar, portanto que o produto resultante do processo de indexação é o índice. Baseados em palavras-chave de qualquer documento, segundo Guinchat e Menou (1983) e Reitz (2007), temos os seguintes tipos de índice:

- KWIC (**Keyword in context**): palavra-chave no contexto, índice permutado de palavras-chave retiradas do título documento, resumo ou por vezes o texto completo. Pode ser realizado manualmente ou automaticamente, com a disposição das palavras-chave extraídas em lista, em ordem alfabética, normalmente centralizada. Cada palavra-chave selecionada de um documento gera uma entrada para RI. Para um documento, portanto, pode-se ter várias formas de acesso a eles, permutando-se o termo em destaque, o que é identificado como um nova entrada.
- KWOC (**Keyword out of context**): palavra-chave fora do contexto, uma variação do KWIC. Palavras-chave do título de um documento e por vezes do corpo do texto são extraídas formando um índice. Nesse índice cada termo é seguido, normalmente, pelos títulos completos dos documentos que as contêm.

Seguem outros tipos de indexação no Quadro 13 abaixo.

Quadro 13. Tipos de indexação

Tipos de Indexação	Características
Associativa	Indexação automática que indica a relação de vizinhança entre dois termos, sem que isso implique em associação funcional específica.
Automática	Efetuada por computador.
Contextual	Utilização das palavras significativas, apresentadas no seu contexto.
Coordenada ou correlativa	Indicação dos assuntos mediante emprego de termos de indexação de igual valor, que podem ser combinados no momento da indexação (pré-coordenação) ou da recuperação (pós-coordenação).
De citações	Relação entre documentos por intermédio das citações comuns entre eles.
De conceitos	Exige a análise do conteúdo temático dos documentos. Tipos de conceitos: entidades, atividades, abstratos, propriedades, heterogêneos.
Derivada	Elaboração de índices baseada no uso de elementos que se encontram no documento.
Em cadeia	Produção de índice alfabético dos termos, ou frases, correspondentes a partes de um índice de classificação.
Hierárquica	Indicação dos registros de informação sob as classes gerais, bem como sob as classes específicas que lhes são subordinadas.
Na fonte	Inclusão no documento, quando da sua editoração, dos termos de indexação necessários à recuperação da informação nele contido.
Por palavra-chave	Utilização das palavras significativas de um texto para indicação dos assuntos.
Por relações ou relacionada	Representação estrutural de assuntos complexos, pela interposição das relações formalmente indicadas entre os termos.
Por truncamento	Técnica automática que efetua a seleção de termos conforme critérios lingüísticos, fundamentados no uso da raiz das palavras.
Por unitermos	Tipos de indexação coordenada que emprega palavras simples.

Fonte: Cavalcanti (1982)⁵⁰, adaptado por Feitosa (2005, p.170).

Pelo destaque no Quadro acima, percebe-se que a indexação não se concentra exclusivamente no conteúdo dos documentos, apesar de ser esse seu principal objetivo. Com sua idéia básica de criação de índices, pode-se afirmar que um conjunto de documentos por ser indexado pelos seus autores, tem-se, portanto, um índice de autores, que pode facilitar a RI. Há também o um índice de citações, destacado no quadro acima.

⁵⁰ CAVALCANTI, C. R. **Indexação**. Brasília: ABDF, 1982. (Estudos Avançados em Biblioteconomia e Ciência da Informação, v.1)

3.5 Metadados

Metadados usualmente são informações sobre um recurso informacional, que descreve um documento. Mendéz Rodríguez (2002, p.41) define metadados no contexto da textualidade eletrônica, como “[...] qualquer dado que faz alusão à identificação, descrição e localização de recursos eletrônicos em rede. São, pois, dados que descrevem os atributos de um recurso, caracterizam suas relações, suporta sua recuperação, gestão e uso efetivo, e existem em um contexto estritamente eletrônico”.

Reitz (2007) afirma que os metadados podem ser categorizados como: descritivo (facilita a indexação, busca, identificação e seleção de documentos); estrutural (descreve a estrutura interna de documentos); e administrativo (ajuda na administração de recursos, podendo incluir metadados técnicos, de propriedade e preservação, que descrevem as características físicas de um recurso).

Buckland (2006) diz que há tipos diferentes de metadados descritivos: o técnico (descrevem formato, padrões de codificação); o administrativo (descreve direitos de propriedade intelectual, condições de acesso); e o de conteúdo (temática, escopo, autoria, etc.). Nessa categorização, o autor considera que todo metadado é descritivo, assim, podemos afirmar, que os metadados são conjuntos de dados que descrevem um recurso informacional, do ponto de vista de conteúdo, de forma, de gestão, etc., sendo um método específico de representação da informação, pois está condicionado aos pacotes informacionais.

Além da função de descrição, Buckland (2006, p.4) afirma que existe outra função para os metadados:

Pensar em metadados como um meio de descrever documentos individuais reflete apenas um dos dois papéis dos metadados. O segundo uso de metadados é diferente: ele ocorre quando você começa com os metadados ao invés dos dados, com a descrição ao invés do documento. Isso ocorre quando você pesquisa em um catálogo ou navega em qualquer índice. Este segundo uso dos metadados é para pesquisa e seleção.

Mas somente o estabelecimento de metadados não garante sucesso da representação e recuperação da informação, principalmente em um ambiente como a Web. Existe a necessidade que eles sejam padronizados, para que haja interoperabilidade. Ou seja, existe a necessidade de um padrão que seja

usado para que os sistemas de informação e os indivíduos que os usam, possam entender a estrutura de qualquer recurso informacional.

Um exemplo de formato de metadados é o MARC (MACHine-Readable Cataloging – Catalogação Legível por Computador), que se trata de “um conjunto de padrões para identificar, armazenar, e comunicar informações bibliográficas em formato legível por máquina, de forma que diferentes computadores e programas possam reconhecer, processar e estabelecer pontos de acesso dos elementos” (ALVES; SOUZA, 2007, p.25). O MARC é um formato comum utilizado em bibliotecas, principalmente as universitárias. O Quadro 14 ilustra a sua estrutura básica.

Quadro 14. Estrutura básica de metadados do Formato MARC

Campos	Descrição
0XX	Informações de controle, números e códigos
1XX	Autoria (nome pessoal, entidade, evento)
2XX	Títulos, edição, imprensa
3XX	Descrição física
4XX	Série
5XX	Notas
6XX	Entradas de assuntos
7XX	Entradas secundárias (nome pessoal, entidade, evento, título)
8XX	Entradas secundárias de série
9XX	Uso local

Fonte: Alves e Souza (2007, p.26).

A partir dos campos, há desdobramentos para especificar ainda mais os níveis de descrição dos recursos informacionais por meio dos metadados do Formato MARC.

Outro padrão de metadados que tem seu foco no ambiente da Web, sendo adotado no projeto da WS, é o *Dublin Core* (DC). Souza, Vendrusculo e Melo (2000, p.93) comentam que:

O conjunto de metadados descrito pelo DC é composto de 15 elementos, os quais poderiam ser descritos como o mais baixo denominador comum para descrição de recurso (equivalente a uma ficha catalográfica) [...] As principais características do padrão DC são a simplicidade na descrição dos recursos, entendimento semântico universal (dos elementos), escopo internacional e extensibilidade (o que permite sua adaptação às necessidades adicionais de descrição).

Os quinze elementos desse padrão são os seguintes, conforme Souza, Vendrúsculo e Melo (2000) e Souza e Alvarenga (2004):

- Título: o nome dado ao recurso informacional (metadado de conteúdo);
- Autor ou criador: a pessoa ou organização responsável pelo conteúdo intelectual de recurso (metadado de propriedade intelectual);
- Assunto(s) ou palavra(s)-chave: termos que descrevem os temas que cobrem o conteúdo intelectual o recurso (metadado de conteúdo);
- Descrição: descrição textual do conteúdo do recurso, resumo para documentos textuais e descrição para recursos visuais (metadado de conteúdo);
- Publicador ou editor: o responsável (indivíduo ou organização) por tornar o recurso ou documento disponível (metadado de propriedade intelectual);
- Contribuidores ou colaboradores: indivíduos ou organizações que contribuíram para a criação do conteúdo do recurso (metadado de propriedade intelectual);
- Data: data em que o recurso foi tornado disponível (metadado administrativo);
- Tipo: categoria do recurso, se imagem, texto, vídeo, etc. (metadado de conteúdo);
- Formato: estabelecido o tipo de recurso, determina-se o seu formato, o tipo texto, pode ser digital ou impresso (metadado administrativo);
- Identificador: número que caracteriza o recurso como único (metadado administrativo);
- Fonte: informação sobre outro recurso do qual o presente recurso é derivado (metadado de conteúdo);
- Linguagem ou idioma: o idioma do conteúdo intelectual do recurso (metadado administrativo e de conteúdo);
- Relação: permite o relacionamento com outros recursos, como, por exemplo, se é um capítulo em um livro (metadado administrativo);
- Cobertura: onde o recurso está fisicamente localizado (metadado administrativo);
- Direitos autorais: declaração de direito sobre a propriedade, um identificador que *link* a uma declaração de direito sobre a propriedade, ou um identificador que *link* a um serviço que fornece informação sobre o direito de propriedade do recurso (metadado de propriedade intelectual).

Souza, Vendrúsculo e Melo (2000, p.95) afirmam que a partir da aplicação do padrão DC em determinada instituição, houve uma adaptação, com o acréscimo de mais três campos, para o uso específico da instituição:

- Categoria: “Atribuição de categoria de assunto que expresse a essência do conteúdo do recurso. O uso de sistemas de classificação formais é encorajado, como o Agrícola, Dewey Decimal Classification” (metadado de conteúdo). Aparece como elemento para classificar o recurso, enquanto o elemento assunto ou palavra-chave é utilizado para indicar a indexação do conteúdo do recurso.
- Acesso: “identificador de recursos capturados da rede, como URLs e URNs (quando implementadas). Há outros identificadores internacionalmente conhecidos, como International Standard Book Number (ISBN)” (metadado administrativo);
- Contato: “indicação para contato de nome e e-mail da pessoa ou instituição à qual o recurso está vinculado” (metadado administrativo).

Esse exemplo de adaptação, indica que o DC é realmente apenas um modelo básico para a organização dos recursos informacionais em ambiente digital. Miranda (2007), em uma das dissertações analisadas, conclui entre diversas coisas, que o padrão de metadados DC não cobre as necessidades entendidas por sua pesquisa, como suficientes para a descrição de imagens.

Através da adoção desse padrão de metadados, a interoperabilidade acontece mais facilmente no ambiente Web. Mas, como afirmado trata-se de uma base, pois a necessidade de fazer um tratamento adequado de outros tipos de informações pode requerer a inclusão de outros campos de informação.

3.6 Tesouros

O termo tesouro é antigo e tem sua origem etimológica na palavra latina *thesaurus*, que por sua vez tem origem no grego *thesaurós*. Para ambas o significado é tesouro ou repositório de palavras

(ARANO, 2005). Ou seja, é a parte mais significativa de um conjunto de termos.

Como já visto, se trata de uma linguagem documentária controlada. Entende-se ser um instrumento de RC, pois auxilia no processo de indexação, mas não tem ligação direta com recursos de informações, ou seja, pode não ser aplicado diretamente à indexação de suportes físicos. Guinchat e Menou (1983, p.148) definem o tesauro como:

Um conjunto controlado de termos entre os quais se estabelece relações hierárquicas e analógicas, termos para os quais, se necessário, se dão as relações de sinonímia com as palavras da linguagem natural, e que se aplica em um campo particular do conhecimento. Um tesauro pode ter desde várias centenas até umas vinte mil expressões; em termo médio se situa aproximadamente entre mil. Os termos tem uma significação única, não ambígua; se eliminam os casos de polissemia y se controla os sinônimos. Os tesauros são geralmente especializados [...]

Slype (1991), afirma serem os tesauros listas estruturadas de conceitos, destinados a representar de maneira unívoca o conteúdo de recursos informacionais e as consultas a um sistema determinado de documentação, como também ajudar o usuário na indexação dos documentos, bem como auxiliar na recuperação da informação.

Sobre a evolução dos tesauros Moreira, Alvarenga e Oliveira (2004) afirmam que essa evolução acontece em duas linhas: uma que tem sua base no sistema unitermo, introduzido por Mortimer Taube em 1951 (americana), motivada pela necessidade de melhorar as limitações de uma linguagem documentária prévia, o Unitermo; a segunda, influenciada pela teoria da classificação facetada (britânica, influenciada por Ranganathan em 1930), aplica o uso de categorias para organização de conceitos em um domínio, que permite uma melhor organização das hierarquias e posicionamento dos termos associados com os conceitos. Ambas as linhas da evolução do tesauro tende a convergir para o padrão ISO 2788, como também padrão BS-5723 (padrão britânico).

Para esses mesmos autores, são dois os aspectos dos tesauros para a CI: o estrutural (o tesauro é um vocabulário dinâmico controlado de termos relacionados semanticamente e coberto por uma relação genérica de um domínio específico do conhecimento); e o funcional (um instrumento de controle terminológico usado na tradução da linguagem natural de documentos, de um indexador ou de usuários em um sistema de linguagem restrita ou controlada).

Em geral se trata de instrumento com relação entre os conceitos, organizado de forma hierárquica,

assim geralmente são acompanhados de índices, para facilitar a consulta.

Como exemplo, apresentamos o sumário do Tesouro “Docutes: tesouro de Ciencias de la Documentación”, no Quadro 15.

Quadro 15. Sumário do ‘Docutes: tesouro de Ciencias de la Documentación’

Introdução (Objetivos, Metodologia e fases de elaboração, Docutes: estrutura e composição, Bibliografia)
Índice alfabético
Índice hierárquico
Índice permutado KWOC
Índice Inglês-Espanhol

Fonte: Adaptado de Alvite Díez et al.(2004).

A partir desse quadro, percebe-se um tesouro pode ser chamado de um índice hierárquico, com foco no controle de adesão dos termos mais significativos. Percebe-se também, que outros índices são importantes para ajudar no uso do instrumento.

No quadro 16 têm-se as classes principais que orientam a construção desse tesouro, chamados pelos autores de campos semânticos (ALVITE BÍEZ et al., 2004).

Quadro 16. Classes principais do ‘Docutes: tesouro de Ciencias de la Documentación’

1 Ciencias de la Documentación: Historia. Teoría. Sistemas.
2 Información. Documentos. Fuentes de Información.
3 Investigación y Metodología documental.
4 Representación y recuperación de la información.
5 Sistemas de información.
6 Tecnologías de la información.

Fonte: Alvite Díez et al.(2004).

Outra característica que ajuda no uso de um tesouro, colaborando principalmente para o estabelecimento das relações e das restrições dos termos, são os seguintes símbolos que seguem, que para o exemplo do presente tesouro estão apresentados nos índices alfabético e permutado KWOC:

- USE: quando o termo aparece com essa indicação, ele não é o adotado, assim o índice remete à outro preferencial. Exemplo:

Atitude do usuário

USE: Comportamento do usuário

- UP (Usado para): A condição é invertida ao USE, o índice diz que tal termo é preterido à outro. Exemplo:

Comportamento do usuário

UP: Atitude do usuário

- TC (Termo categoria ou Termo principal): são os campos semânticos que dividem o tesouro, mas não deixam de ser o termo geral. Exemplo:

Informação

TC: 03. Informação. Documentos. Fontes de informação

- TG (Termo geral): indica termo imediatamente superior na hierarquia do tesouro, conforme mostra o exemplo.

Informação pública

TG: Informação

- TE (Termo específico): indica termo imediatamente inferior, sempre associado a um termo geral (TG), conforme mostra o exemplo.

Informação

TE: Acesso à informação

Aspectos éticos

Aspectos legais

Aspectos psicológicos

Aspectos socioculturais

Barreiras lingüísticas

- TR (Termo relacionado): relação semântica horizontal.

Informação

TR: Atualidade dos documentos

Alfabetização informacional

Propriedade intelectual

- NA (nota de aplicação): esclarece o sentido de uso do termo.

Livros antigos

NA: Livros publicados antes de 1801

Segue abaixo o Quadro 17, que apresenta o exemplo do tesauro, a partir do índice hierárquico, que demonstra as relações conceituais.

Quadro 17. Parte do índice hierárquico do ‘Docutes: tesouro de Ciencias de la Documentación’

02. Información.Documentos.Fuentes de información
02.01 . Documentos
.. Contenidos documentales
.. Documentos administrativos
... Ciclo vital del documento
.... Etapa activa
.... Etapa inactiva
.... Etapa semiactiva
... Documentos de constancia
... Documentos de decisión
... Documentos de iniciación
... Documentos de instrucción
... Documentos de juicio
... Documentos de los administrados
... Documentos de terminación
... Documentos de transmisión
... Documentos endógenos
... Documentos exógenos
... Documentos resolutorios
.. Documentos audiovisuales
... Películas
... Vídeos
.. Documentos bibliográficos
... Libros
.... Libro fácil

Fonte: Alvite Díez et al.(2004, p.219).

Os tesouros podem trazer contribuições para a OI na Web, mesmo por que são instrumentos de RC estabilizados, que lidam com áreas específicas do conhecimento.

3.7 Taxonomias

Taxonomia, de forma ampla é a ciência da classificação. Reitz (2007) diz que é um termo tradicionalmente usado na ciência da vida para classificar organismos vivos e todas suas variações. Atualmente tem sido aplicado na Web, para a classificação de recursos digitais.

Zhonghong, Chaudhry e Khoo (2006) afirmam que a taxonomia difere da classificação e dos tesauros em diversos aspectos, como mostrado no Quadro 18 abaixo.

Quadro 18. Diferenças entre taxonomias, sistemas de classificação e tesauros

CARACTERÍSTICAS		SISTEMAS DE CLASSIFICAÇÃO	TESAUROS	TAXONOMIAS
Escopo		Comunidade de bibliotecários Disciplinas acadêmicas	Ambiente digital Comunidade acadêmica	Ambiente de interface na Web Ambiente organizacional
Objetos Tratados		Coleções	Documentos	Recursos digitais
Funções		Classificação Arquivamento	Indexação Pesquisa	Categorização Busca e navegação
Formas	Estrutura hierárquica	Unidimensional Uso combinado de notações	Relações conceituais	Estrutura dinâmica
	Termos	Classes	Termos	Categorias
Foco		Mais em conteúdo	Mais em conteúdo	Mais em usuários

Fonte: Adaptação de Zhonghong, Chaudhry e Khoo (2006, p.166).

A taxonomia não deixa de ser um tipo de classificação, mas ela vai além, sendo altamente contextual, organizacional e dinâmica, ou seja, é uma atividade RC, não possuindo relação direta com os documentos. Apesar de serem aplicadas nos ambientes digitais, elas objetivam ajudar os usuários na navegação no ambiente Web, servindo como uma espécie de mapa do conhecimento, não sendo tão básicas como os sistemas de classificação, nem tão exaustivas quanto um tesauro, uma vez, que esse último, geralmente é objeto que auxilia a indexação, sendo utilizado principalmente por especialistas para auxílio na RI.

3.8 Ontologias

O termo ontologia usado pela computação tem pouca relação com o termo usado na filosofia. Para a CC a ontologia busca estruturar informação a partir de conceitos e suas relações. Outros exemplos dessa forma de estruturação da informação são: tesauros e redes semânticas. As formas anteriores

para estruturação da informação, a partir da utilização de termos, classificação e criação de categorias (ALMEIDA; BAX, 2003), já não se mostravam suficientes, apesar de importantes e em uso, no que diz respeito ao ambiente Web.

Moreira, Alvarenga e Oliveira (2004) dizem que o termo ontologia começou a ser usado na CC, na subárea da IA, no início dos anos 90 do século XX, em projetos cuja meta era organizar grandes bases de conhecimento para compartilhamento e reuso, na busca de diminuir custos. Mas esse objetivo só poderia acontecer se existir um entendimento claro dos acordos ontológicos associados. Tais acordos determinam o que é relevante em um certo domínio, então isto é representado em uma base de conhecimento. O compartilhamento dos acordos ontológicos torna possível a comunicação entre agentes (humanos ou não), para estabelecimento da compreensão comum de um domínio específico do conhecimento.

Assim, as ontologias permitem o estabelecimento de regras para um programa deduzir significados da informação em um recurso informacional, ou seja, elas possibilitam manipular os termos de uma maneira mais útil e eficiente, envolvendo a definição de uma hierarquia de classes e subclasses para os conceitos envolvidos (MORAIS; SOARES, 2002, p.5).

Uma ontologia objetiva possibilitar estrutura semântica para esquemas de metadados, facilitando a comunicação entre sistemas e agentes, ou seja, tem seu foco nas máquinas. Assim, ontologias estabelecem fundamentos de significados conceituais, sem os quais a construção e o uso da informação na Web continuariam sendo destinados aos usuários humanos (JACOB, 2003), impossibilitando a implantação da WS.

Segundo Moreira, Alvarenga e Oliveira (2004), o uso dessa ferramenta pode possibilitar benefícios para: representação do conhecimento, reuso do conhecimento, compartilhamento do conhecimento, aquisição do conhecimento e integração do conhecimento, processamento da linguagem natural, tradução automática, troca de informação entre sistemas, agentes, companhias ou pessoas.

Sobre essa afirmação anterior, entende-se haver um conflito de significação dos conceitos conhecimento e informação, pois reuso, compartilhamento, aquisição e integração em um sentido restrito, estão ligados à informação enquanto coisa, enquanto conhecimento objetivo.

Souza e Alvarenga (2004, p.137) afirmam que “as ontologias se apresentam como um modelo de relacionamento de entidades e suas interações, em algum domínio particular do conhecimento ou específico a alguma atividade”.

Apesar de ser um assunto muito discutido na literatura, não há consenso sobre seu objetivo e sua definição. Moreira, Alvarenga e Oliveira (2004) apresentam quatro interpretações sobre as ontologias discutidas na literatura:

- Como um sistema conceitual informal, abaixo de uma base de conhecimento particular (nesse caso ontologia pertence ao nível conceitual e não ao nível simbólico);
- Como uma representação de um sistema conceitual via uma teoria lógica (define ontologia como um tipo de base especial do conhecimento no nível simbólico; conhecimento independente de um estado de domínio particular);
- Como um vocabulário usado por uma teoria lógica (o nível de formalização desse vocabulário pode variar de uma ontologia a outra); e
- Como uma especificação de conceitualização (uma conceitualização pode ser entendida como um conjunto de comitês ontológicos – nível simbólico).

Esses mesmos autores realizaram um estudo para diferenciar os objetivos dos tesauros e das ontologias. Eles afirmam que nas definições de ontologia encontradas na literatura o termo “formal” e “lógica” estão presentes, revelando uma certa preocupação com o rigor das representações, uma vez que a ontologia é usada para a troca de informações entre sistemas computacionais. O tesauro também pressupõe a ausência de ambigüidades na definição dos termos; contudo, essa condição é assegurada pela construção da metodologia. A ontologia enfatiza a formalização das propriedades do grupo de relacionamento necessária para acoplar inferências através de um computador, enquanto que no tesauro a inferência é acoplada pelas intervenções humanas. A necessidade da inferência, no caso da ontologia, tem suas raízes na troca de informação automática entre sistemas de computador (MOREIRA; ALVARENGA; OLIVEIRA, 2004).

As ontologias apresentam semelhança com as classificações, com os tesauros e com as taxonomias, pois tratam de: vocabulário, conceito, modelo de relacionamento de entidades, etc. O Quadro abaixo, desenvolvido a partir do Quadro 18, enfatiza a ontologia, procurando estabelecer as diferenças entre ela e as taxonomias, os tesauros e os sistemas de classificação.

Quadro 21. Diferenças entre taxonomias, sistemas de classificação, tesouros e ontologias.

CARACTERÍSTICAS		Ver Quadro 20, página 133.	ONTOLOGIA
Escopo			Ambiente estrutural e lógico na Web
Objetos Tratados			Recursos digitais
Funções			Possibilitar inferências pelos softwares inteligentes
Formas	Estrutura hierárquica		Relações conceituais Estrutura dinâmica
	Termos		Conceitos
Foco			Mais em softwares inteligentes

Concluem Moreira, Alvarenga e Oliveira (2004), que os tesouros na CI e as ontologias usadas na CC possuem diferentes origens e propostas, os primeiros nasceram como um instrumento prático para socorrer indexadores e pesquisas de documentos, e mais tarde da necessidade de descrever objetos e suas relações. Na CC, a situação ainda não é clara, ou seja, tudo que modela um segmento da realidade pode ser chamado de ontologia, sendo assim, mesmo os tesouros podem ser estruturados como uma ontologia terminológica.

4 TECNOLOGIA E SOCIEDADE

A Internet e, principalmente a Web são as principais tecnologias estudadas nesse trabalho. A relação delas com a sociedade é nítida e cresce a cada dia, no âmbito da chamada cibercultura. No entanto, parece necessário responder a algumas questões preliminares, como: o que é a tecnologia, como sua relação com a sociedade é pensada, devido aos benefícios e também a prováveis problemas. Assim, apresenta-se em um primeiro momento a relação da tecnologia, de forma geral, com a sociedade e depois foi discutida a cibernética, primeiro passo para a automação de diversas atividades do homem, inclusive o pensamento. Em seguida, foram apresentadas a Internet e a Web, seus conceitos básicos, suas diferenças, seus históricos, personalidades envolvidas.

4.1 Algumas reflexões sobre o impacto da tecnologia

Há quem entenda a tecnologia como “o campo de conhecimento relativo ao desenho de artefatos e à planificação da sua realização, operação, ajuste, manutenção e monitoramento à luz do conhecimento científico. Ou, resumidamente, o estudo científico do artificial” (BUNGE, 1985, p.231⁵¹ citado por CUPANI, 2004, p.496).

Cupani (2004, p.499) apresenta a abordagem da tecnologia sob o enfoque fenomenológico de Borgmann⁵², em que “a tecnologia é o modo tipicamente moderno de o homem lidar com o mundo, um ‘paradigma’ ou ‘padrão’ característico e limitador da existência, intrínseco à vida quotidiana”. Ou seja, os avanços tecnológicos não modificam apenas as atividades exteriores do homem, provocam também alterações profundas no interior de cada indivíduo. Assim, parecem necessárias algumas reflexões sobre como a tecnologia se relaciona com a sociedade, já que ela se insere em todas as suas atividades, direta ou indiretamente, seja na saúde, educação, habitação, lazer e entretenimento, etc., isto é, na necessidade de um medicamento, no acesso a um curso, no acesso ao domicílio (elevador), ao ouvir música, na comunicação, etc.

⁵¹ BUNGE, M. **Treatise on basic philosophy**. Dordrecht: Reidel, 1985. Tomo 7: Philosophy of science and technology.

⁵² Albert Borgmann, professor da Universidade de Montana (EUA) e autor de outras obras como *Crossing the postmodern divide* (1992) e *Holding on to reality: the nature of information and the turn of the millennium* (1999).

Schwartz⁵³ (1975, p.48) citado por Moraes (1988) afirma que “poderíamos viver sem muitas das falsas necessidades modernas como a eletricidade, televisão, automóveis, etc., mas acharíamos a vida quase intolerável sem os alimentos, o fogo, os instrumentos simples, a agricultura, a linguagem”. Essa dificuldade de viver sem os avanços tecnológicos pode ser ainda mais acentuada se pensarmos nos *nativos tecnológicos*⁵⁴, pessoas que conheceram o mundo e vivenciam suas relações sociais sempre e já nessa perspectiva, diferente dos indivíduos que podemos chamar de *imigrantes tecnológicos*⁵⁵, pessoas que vivem nessa perspectiva tecnológica atual, mas são de uma geração não tão dependente, os indivíduos da máquina de escrever, da escassa telefonia fixa, da fotografia analógica, ou indo mais longe, do fogão à lenha, do carro de boi, da lamparina à querosene, da parteira, da cuia de madeira.

O objetivo da técnica, teoricamente, é o de humanizar a natureza, ou seja, transformá-la em uma parte do homem, para seu auxílio e bem-estar. Pode-se afirmar que atualmente o desenvolvimento tecnológico não tem apenas esse objetivo. Isso é evidente não só pelo controle da natureza, mas também pela sua degradação, além dos interesses econômicos (MORAIS, 1988).

Moraes (1988) apresenta algumas reflexões sobre o que ele chama de civilização pré-tecnológica e tecnológica, pois ainda convive-se com uma geração que vivenciou ambos os períodos, e parece necessário entender essas análises. Os principais pontos destacados pelo autor são: 1) o totalitarismo que se insinua pela técnica e 2) a diminuição das diferenças entre as sociedades, ou seja, ocorre uma *globalização* de quase tudo: cultura (música, artes plásticas, literatura, etc.), economia, consumo, etc. Certamente há muitos benefícios, mas há necessidade de pensarmos criticamente todo esse desenvolvimento, olhando para a tecnologia como uma forma, um meio, e não um fim.

Voltamos a afirmar que uma contra-tecnologia ser-nos-ia, hoje, impossível. O que há é a necessidade de que reinterpretemos a tecnologia, recolocando o *capital não-vivo* a serviço do *capital-vivo*; em outras palavras, há uma urgência de que neguemos a tecnologia como um *fim*, e a recoliquemos como *meio* de afirmação do humano (MORAIS, 1988, p.111, grifos do autor).

Para entendermos melhor essa dinâmica da civilização tecnológica, Moraes (1988) propõe algumas subdivisões:

⁵³ SCHWARTZ, E. S. **A inflação da técnica**. São Paulo: Melhoramentos, 1975. 312p. (Hoje e Amanhã)

⁵⁴ Termo utilizado por Aldo de Albuquerque Barreto, quando da palestra intitulada “A diversidade cultural, a estrutura e o fluxo da informação”, no encerramento do IX Encontro Nacional de Pesquisa em Ciência da Informação (IX ENANCIB), organizado pela Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), em São Paulo, 2008.

⁵⁵ Idem anterior.

- Princípios que a regem. Fromm (1965)⁵⁶ citado por Moraes (1988) aponta três princípios principais que balizam a civilização tecnológica: a) tudo o que é tecnicamente possível de fazer-se, deve ser feito, ou seja, temos aqui um princípio que pode fazer com que o desenvolvimento tecnológico não ocorra pela necessidade da sociedade, mas uma sociedade que se desenvolva a partir do que a tecnologia oferece; b) o atual avanço tecnológico deve conduzir ao ideal de absoluta eficiência, isto leva às seguintes questões, o que é a eficiência, e como tal benefício será de todos; c) quanto mais produzir melhor, isto é, o detrimento da qualidade diante da quantidade.
- As “*quase-soluções*” tecnológicas. Evidenciadas pelo surgimento de novos problemas, que podem aparecer das incompletas soluções propostas, do aumento do problema original e de efeitos secundários da solução originalmente proposta (SCHWARTZ, 1975⁵⁷, citado por MORAIS, 1988).

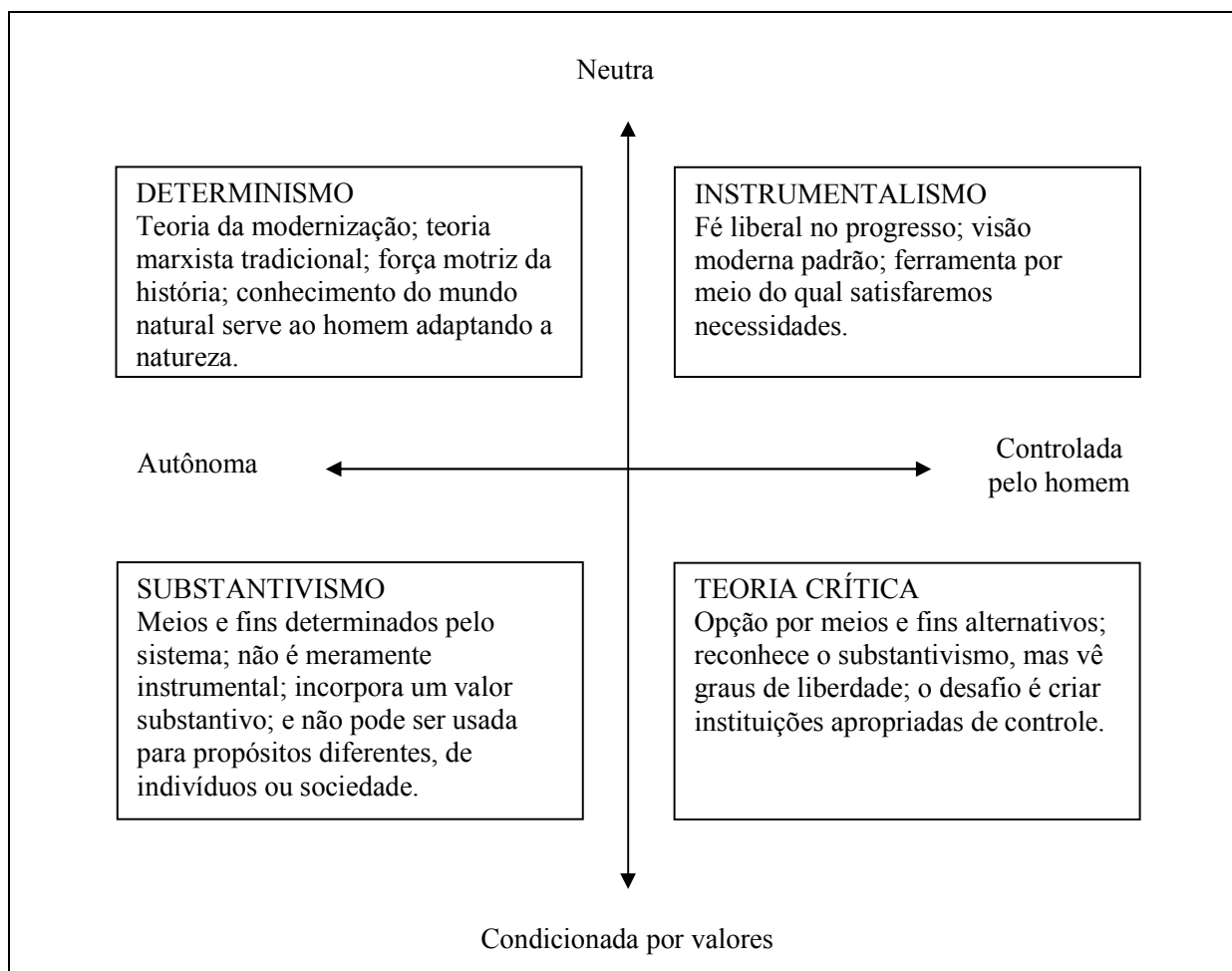
Como Moraes (1988) salientou, o caminho para o qual a tecnologia levou a sociedade parece não ter volta, mesmo porque há muitos benefícios e interesses que sustentam essa dependência social das técnicas. Surge, porém, a emergência de se pensar toda essa influência, e principalmente de estender os benefícios tecnológicos a todos, pois, a partir do que o autor acima denomina de “*quase soluções*” tecnológicas surgem diversos outros problemas.

Vejamos como exemplo o caso da Web, que ainda não tem acesso oferecido gratuitamente a toda população. Ela contém uma série de fontes de informações exclusivas, ou seja, que só podem ser acessadas através dela, mas se trata de uma tecnologia que não é de alcance de todos, não apenas pela barreira financeira, também pela tecnológica (há a necessidade de equipamento adequado para o acesso à Web) e pela barreira educacional (há a necessidade de conhecimento para a operação dos equipamentos e da própria Web). Isto é, tem-se uma “*quase-solução*” tecnológica que aparece para solucionar alguns problemas, mas acaba por gerar outros.

Para Dagnino, Brandão e Novaes (2004), as visões da tecnologia são quatro, conforme indica a Figura 9 abaixo.

⁵⁶ FROMM, E. **A revolução da esperança**. Rio de Janeiro: Zahar, 1965. 169p.

⁵⁷ Idem 50.

Figura 9. Quatro visões sobre a tecnologia.

Fonte: Dagnino, Brandão e Novaes (2004, p.48).

Controlada pelo homem e com neutralidade de valores, tem-se o Instrumentalismo, que se trata de uma visão moderna, concebendo a tecnologia como um instrumento ou ferramenta que o homem utiliza para satisfazer suas necessidades, direcionando seu desenvolvimento de acordo com sua vontade. No Determinismo, tem-se a concepção de uma tecnologia autônoma e neutra, onde o desenvolvimento tecnológico é a força da história, ocorrendo sem o controle do homem e moldando a sociedade segundo suas necessidades de eficiência e progresso. A terceira visão, o Substantivismo, entende a tecnologia dotada de valores, mas autônoma, ou seja, respectivamente, não neutra e sem o controle do homem. Na última visão, a tecnologia é concebida pela Teoria Crítica, sendo, portanto, controlada pelo homem e condicionada por valores (tendenciosa). Essa visão reconhece as consequências do desenvolvimento tecnológico como desastrosas (substantivismo), mas considera que ainda existe possibilidade de liberdade pela técnica, vê, observa, como principal problema, o

controle humano sobre a tecnologia (DAGNINO; BRANDÃO; NOVAES, 2004, p.48).

Luís Garcia (2007, p.328) tece algumas reflexões sobre a obra de Simmel⁵⁸, destacando três elementos para uma reflexão sobre a tecnologia. Em primeiro lugar, “a tecnologia resulta da exteriorização do espírito subjetivo no mundo social e cultural e nessa qualidade de objetivação da subjetividade humana integra, por conseguinte, plenamente o processo cultural”. Assim, complementa o autor “o que caracteriza o mundo moderno como tecnológico não decorre simplesmente do desenvolvimento material extensivo, mas também da disseminação deste fenômeno para outras áreas da vida apenas formalmente afastadas do mesmo”. Ou seja, a tecnologia definitivamente está inserida na maneira de ser sociedade e dessa forma ela mais influencia do que é influenciada e esse é o segundo elemento apresentado pelo autor, a tecnologia é intrínseca na relação do homem com o mundo. Em terceiro lugar,

[...] como resultado da evolução dos meios e do fenômeno da divisão do trabalho associado à diferenciação capitalista e ao estímulo do dinheiro, a tecnologia adquire o caráter de sistema autônomo, ilimitado e universal, no sentido de fim último e enganoso da ação e do desenvolvimento humanos. Trata-se da alienação do homem em relação aos seus próprios instrumentos e artefatos: a tecnologia deixa de funcionar em relação aos fins do homem; a humanidade passa a funcionar em função da tecnologia (LUÍS GARCIA, 2007, p.328).

A posição apresentada por Luís Garcia (2007) difere da de Dagnino, Brandão e Novaes (2004), que acreditam que a tecnologia é tendenciosa (condicionada por valores), mas se controlada pelo homem pode ter algo de sua proposta inicial, ou seja, ser benéfica para a sociedade, sendo desenvolvida a partir de suas necessidades. O primeiro acredita na autonomia da tecnologia, condicionada por valores, ou seja, ela imprime suas condições, como um organismo “vivo”.

Pois bem, parece recomendável que a sociedade pense e reflita sobre a influência da tecnologia, sobre os interesses existentes por trás do seu desenvolvimento, principalmente pelos limites das necessidades sociais. Mattelart (2006, p.156) afirma que diante da constatação de carências que ocorrem na sociedade, “os discursos promocionais sobre a eminência da salvação pela tecnologia beira a indecência”.

⁵⁸ Georg Simmel (1858-1918) foi um sociólogo alemão.

4.2 A automação e os períodos da computação

Morais (1988) nas suas reflexões sobre essa relação conflituosa de tecnologia e sociedade destaca a Cibernética e a Automação, para exaltar os benefícios da tecnologia em geral, indicando que essas soluções propostas se apresentam como ferramentas que ampliam os sentidos do homem. Mas há de se considerar que Cibernética não se apresenta apenas em relação aos computadores. Ela surge da necessidade de analisar erros e indeterminações presentes nas atividades humanas, principalmente na ciência e na tecnologia; ela surge para dominar a dúvida. D’Azevedo (1972, p. 11)⁵⁹ citado por Moraes (1988, p.136) conceitua a Cibernética como “a ciência ou arte do comando e do controle de um processo organizado, com a capacidade de autocorreção e realimentação próprias, que lhe imprimem o máximo de eficiência”, ou seja, é o estudo dos sistemas, sejam eles humanos ou artificiais. A Cibernética traz consigo os seguintes conceitos para anunciar a idéia de sistema, que fundamentam o surgimento do computador:

- Entropia: tudo no universo tende a se desorganizar naturalmente. Assim Wiener (1970, p.32)⁶⁰ citado por Moraes (1988, p.138) diz que “as máquinas simuladoras de inteligência humana foram concebidas para desempenhar a missão de bolsões de entropia decrescente dentro de um mundo em que a entropia geral tende a aumentar”. O Quadro 20 apresenta a idéia da entropia.

Quadro 20. A entropia como fundamento da Cibernética

ENTROPIA	Situação: Incerteza	Situação: Comando	REORGANIZAÇÃO e MANUTENÇÃO
	a) descontrole b) desordem c) incorreção d) ineficiência	a) controle b) organização c) autocorreção e) eficiência máxima	

Fonte: Moraes (1988, p.17).

A entropia como fundamento da cibernética indica a necessidade de preparação por parte de um sistema artificial de se prevenir contra o erro, contra a incerteza. Um sistema, seja humano ou artificial, se vê por vezes diante de uma situação que não é comum: a entropia. Nessa situação, surge

⁵⁹ D’AZEVEDO, M. C. **Cibernética e vida**. Petrópolis: Vozes, 1972. 146p.

⁶⁰ WIENER, N. **Cibernética e sociedade**. 3.ed. São Paulo: Cultrix, 1970. 190p.

a desordem, o descontrole. Assim, tal sistema deve estar preparado para responder a essa situação, e utilizar a informação gerada para evitar que ela aconteça novamente.

- Retroalimentação: “equipamentos executam a reinformação (retroalimentação) trazendo de novo à entrada para processamento os resultados da operação anterior como novos dados” (MORAIS, 1988, p.140), ou seja, o *feedback*.
- Homeostase: quando o sistema “é capaz de manter algumas de suas variáveis dentro de certos limites especificáveis” (APTER, 1973, p.36⁶¹ citado por MORAIS, 1988, p.141).

A partir dessas noções básicas são concebidos os equipamentos como os computadores, formados por: “a) Entrada de informações – INPUT; b) Armazenamento ou memória – STORAGE; c) Saída de resultados do processamento – OUTPUT, que devem estar acoplados a *servomecanismos* que providenciam a retroalimentação auto-corretiva” (MORAIS, 1988, p.143).

Assim, “[...] o computador é um meio artificial de pensamento, criado que foi para ser o apêndice mais veloz da mente humana”, realizando “em pouco tempo cálculos e raciocínios que especialistas humanos deveriam levar centenas de anos para efetuar” (MORAIS, 1988, p.143), ou seja, ele complementa, mas nunca substitui inteiramente o cérebro.

O Quadro 21 abaixo apresenta uma comparação do cérebro humano com o computador.

Quadro 21. Comparação do computador com o cérebro humano

SEMELHANÇAS	DIFERENÇAS
Ambos são dispositivos de uso geral, ou seja, usa-se para diversas atividades.	“Equipotencialidade” do cérebro humano, ou seja, se certas partes do cérebro forem retiradas, outras automaticamente assumem suas funções.
Ambos processam informação.	O computador não é criativo, no sentido radical da palavra.
Ambos incorporam modelos.	Ao lidar com formas, o cérebro humano o reconstrói na falta de algum elemento, o que se chama de função <i>gestáltica</i> . Por enquanto, os computadores não realizam essa atividade.
Chegam à complexidade intelectual realizando grande número de operações simples.	O computador recebe as informações de forma passiva, enquanto o cérebro humano pergunta e argumenta.

Fonte: Apter (1973)⁶² citado por Morais (1988).

⁶¹ APTER, M. J. *Cibernética e psicologia*. Petrópolis: Vozes, 1973. 201p.

Fica evidente, portanto que, as atividades de automação de certa forma, principalmente as voltadas aos computadores, buscaram uma analogia com o cérebro humano, na tentativa de criar seus comportamentos. Podemos afirmar que a automação foi uma forma da Cibernética se estabelecer, indo além de propostas teóricas.

Nas últimas seis décadas (da criação da cibernética em diante), desde que as válvulas substituíram os relés⁶³ dando origem à primeira geração de computadores, a computação vem em intervalos de tempo, cada vez menores, experimentando novos períodos que se consolidam ao longo do tempo. São eles:

1º período: **Computação *Batch* ou processamento em lotes**. Neste modelo computacional predominam os computadores de grande porte (mainframes). O acesso a esses computadores era restrito aos profissionais da área de computação, que utilizavam painéis ou terminais do tipo máquina de escrever, também conhecidos por “console”, para dar instruções ou operar o computador. Nesse ambiente, a interface para entrada de dados para processamento eram os cartões perfurados, e a de saída os relatórios impressos. O usuário final não tinha acesso direto ao computador, ou seja, o processamento de suas transações era efetuado em lotes de cartões perfurados, cada qual referente a um serviço. Lévy (2003) afirma que a informática nesse período servia aos cálculos científicos, ao Estado e às grandes empresas nas estatísticas ou a tarefas pesadas de gerenciamento, como folhas de pagamento. O processamento *batch* ou em lotes, segundo Meirelles (1994, p. 199), é:

[...] execução de programas com pouca ou nenhuma interação com o usuário durante o processamento. Em *batch*, uma vez iniciado um serviço, ele se processa de maneira contínua até o término. O usuário precisa fornecer previamente todas as informações necessárias ao programa e não tem acesso a ele durante o processamento. O processamento é efetuado em série, de forma que um programa só pode ser iniciado após executar o lote anterior, ou seja, trata-se de um processamento seqüencial.

2º período: **Computação *time sharing* ou compartilhada**. Neste modelo computacional, onde ainda predominam os *mainframes*, um ponto forte é o surgimento das primeiras redes de computadores proprietárias⁶⁴, onde o acesso ao núcleo de processamento dessa rede, que são os

⁶² Idem anterior.

⁶³ Dispositivo por meio do qual um circuito é controlado por variações das condições elétricas nele mesmo, ou noutro circuito (FERREIRA, 2004).

⁶⁴ Redes proprietárias são, conceitualmente, redes constituídas por hardware e software de um único fabricante.

mainframes, é efetuado através de terminais “burros”. Em relação a esse modelo computacional, Meirelles (1994, p. 200, grifo do autor) destaca que,

Nessa estrutura de **tempo compartilhado – time sharing** –, o sistema multiusuário tem o computador central funcionando brevemente para cada um dos diversos usuários que têm a sensação de estar o computador à sua disposição, não percebendo que diversos trabalhos estão sendo executados “simultaneamente”. A arquitetura tradicional é um computador central e vários terminais, chamados de **terminais burros** por não terem capacidade de processamento próprio.

Nesse modelo tem-se início a interação humano-computador (umas das subáreas da CC), através de uma interface visual e passiva.

3º período: **Computação pessoal**. Configura-se no início da década de 1980 com o surgimento dos computadores pessoais ou terminais inteligentes, que rapidamente alcançaram o ambiente empresarial, devido à sua capacidade de processamento, armazenamento e outros recursos. Após um período de aceitação e maturação da tecnologia dos computadores pessoais, e com a evolução das TICs, principalmente a dos sistemas operacionais baseados no conceito de janelas (Windows), surgem as primeiras necessidades de interligação desses computadores em rede, inicialmente para compartilhamento de recursos (MEIRELLES, 1998). Os computadores “escapam” do Estado, dos profissionais e pesquisadores da computação e das grandes empresas para segundo Lévy (2000), se tornarem instrumentos de criação (texto), de organização (planilha de dados) e de diversão (jogos) para as pessoas dos países desenvolvidos.

4º período: **Computação em rede**. As redes de computadores ganharam um impulso significativo nas décadas de 80 e 90, e continuaram em crescente desenvolvimento até os dias atuais, passando de redes de compartilhamento de recursos para redes de comunicação, que de acordo com Meirelles (1998, p. 206, grifo do autor) “[...] são formas de interligação entre sistemas de computação que permitem a troca de informações entre eles, tanto em tempo real (on-line) como para troca de mensagens por meio de um disco comum. Essa função é chamada também de **correio eletrônico** [...]”. Os computadores pessoais, uma vez interligados em rede aumentam a capacidade de processamento e de comunicação da rede, e permitem que usuários acessem não somente mainframes, mas principalmente outros computadores centrais, denominados servidores, instalados em redes locais e remotas (distantes). Portanto, tem-se muitas pessoas utilizando computadores em redes públicas e privadas, locais e remotas, através de conexões físicas guiadas (via cabo de rede)

ou não guiadas (através de microondas, rádio frequência, infravermelho e outros meios que não se utilizam de cabos). Apesar da presença dos servidores, o que caracteriza esse período das redes, é que o computador já não o centro, ele é apenas um ponto, um nó (LÉVY, 2000), que quase não tem sentido se não estiver ligado à rede.

A evolução desses períodos culmina no momento atual, onde cada vez mais as pessoas se conectam entre si através de dispositivos computadorizados inteligentes, fixos ou móveis, dentro de uma imensa infra-estrutura de rede de computadores, que é a Internet. O desenvolvimento de aplicativos e ferramentas voltados para uso na Internet têm avançado e produzido importantes resultados devido aos significativos avanços nas áreas de TICs, e atualmente na área da Web. A Internet, desde sua criação, vem se incorporando no cotidiano das pessoas e transformando os computadores pessoais em verdadeiros eletrodomésticos. Da sua influência e da Web surge o chamado o ciberespaço, que Lévy (2000, p.32) afirma ser um “novo espaço de comunicação, de sociabilidade, de organização e de transação, mas também novo mercado da informação e do conhecimento”. O autor (2000, p.92) define o ciberespaço como:

[...] o espaço de comunicação aberto pela interconexão mundial dos computadores e das memórias dos computadores. Essa definição inclui o conjunto dos sistemas de comunicação eletrônicos (aí incluídos os conjuntos de redes hertzianas e telefônicas clássicas), na medida em que transmitem informações provenientes de fontes digitais ou destinadas à digitalização. Insisto na codificação digital, pois ela condiciona o caráter plástico, fluido, calculável com precisão e tratável em tempo real, hipertextual, interativo e, resumindo, virtual da informação que é, parece-me, a marca distinta do ciberespaço. [...] A perspectiva da digitalização geral das informações provavelmente tornará o ciberespaço o principal canal de comunicação e suporte de memória da humanidade a partir do início do próximo século.

A informação, segundo Lévy (2000), em detrimento e desconstrução do computador (hardware e software) se torna o centro desse período, que o autor chama de informática contemporânea. Entende-se, portanto, que esse detrimento e desconstrução do computador a que se refere o autor, faz com que ele já não esteja adequado para servir de analogia ante o cérebro humano. Ele se torna nesse período da informática, analogicamente falando, apenas um neurônio, ou seja, ele agora é apenas um pequeno pedaço do ciberespaço.

Parece, portanto, que a relação da tecnologia com sociedade, no que diz respeito aos dias atuais se estabelece principalmente pelas TICs. Essa relação se firma, criando através de sua infra-estrutura (ciberespaço), o que Lévy (2000) chama de cibercultura.

A cibercultura possui três programas: a interconexão (o isolamento não é inerente à Internet e a Web); as comunidades virtuais (criadas a partir de interesses semelhantes de usuários da rede, independentemente de localização geográfica e relação institucional); e, a inteligência coletiva (a finalidade última e principal da cibercultura) (LEVY, 2000).

Mattelart (2006, p.152) afirma o seguinte sobre esse período:

A integração e a ubiquidade são as palavras-chave. A transgressão das fronteiras é o seu corolário, quer sejam físicas ou funcionais. O local, o nacional e o global se interpenetram. A concepção, a produção e a comercialização são pensadas de modo sincrônico. O recipiente e o conteúdo, o hardware e o software se englobam.

A inteligência coletiva, segundo Lévy (2000), se apresenta como um espaço mais de problemas do que de soluções. Mattelart (2006) diz que surgem nesse período os *inforicos* e o *infopobres*, ou seja, a cibercultura parece estar levando consigo a herança de desigualdades que a sociedade possui em seus diversos setores.

Certamente sabe-se do potencial e dos benefícios das postas da cibercultura, mas questões surgem, ou seja, qual o caminho deve ser seguido? Qual o melhor modelo? Será justo subordinar a sociedade a uma tecnologia que praticamente monopoliza a informação, a comunicação e o conhecimento? A Internet e a Web são pilares desses conceitos, são comumente utilizados como sinônimos, embora não sejam. Segue uma explanação dessas duas tecnologias.

4.3 Internet

O termo Internet vem de *Internetworking* (ligação entre redes), ou seja, na verdade a Internet é um conjunto de redes e *gateways*⁶⁵ que utilizam os protocolos TCP/IP⁶⁶, assim, ela é um conjunto de aparatos físicos (computadores, roteadores, cabos, etc.) e programas (protocolo TCP/IP) utilizados para o transporte de informação (COSTA, 2003).

⁶⁵ Porta de ligação entre protocolos diferentes, permite que os usuários que trabalham em uma rede possam acessar outra (ASSOCIAÇÃO..., 2007).

⁶⁶ “Conjunto dos protocolos de comunicação usados na Internet para gerir a circulação de dados na rede, fragmentando a informação na origem sob a forma de pacotes de dados e reunindo-a novamente no destino, assim como controlando eventuais erros de transmissão” (ASSOCIAÇÃO..., 2007, p.82).

Entre os eventos que parecem interessantes para o desenvolvimento da Internet, destaca-se a criação da Agência de Projetos de Pesquisas Avançadas (*Advanced Research Projects Agency – ARPA*) em 1962, que fez parte do Departamento de Defesa dos Estados Unidos. No mesmo ano, Joseph Carl Robnett Licklider do *Massachusetts Institute of Technology* (MIT) publicou uma série de artigos em que pela primeira vez foi elaborada a idéia da comunicação entre usuários de computadores, sendo contratado pela ARPA em outubro de 1962 (ROBREDO, 2005).

Paul Baram da *RAND Corporation*, entre 1962 e 1964, publicou alguns textos, a partir de um contrato entre a RAND e ARPA, com duas idéias principais: “o uso de uma rede descentralizada com múltiplos caminhos entre dois pontos e a fragmentação das mensagens completas em ‘pacotes’ que ele chamou de ‘blocos de mensagens’, antes de transmiti-los pela rede”. Houve ainda iniciativas de Leonard Kleinrock e Donald Davis, nessa linha de desenvolvimento de sistemas para a transmissão de mensagens por uma rede. Assim um plano completo para a futura Internet, com o nome de ARPANET foi elaborado em 1968 (ROBREDO, 2005, p.243). Em 1969, a rede foi utilizada (LEÃO, 2001;ROBREDO, 2005), servindo a partir de então para que os pesquisadores dispusessem de uma base experimental para o desenvolvimento da Internet.

A Web é apenas mais um serviço da Internet, juntamente com outros principais listados a seguir:

- *FTP (File Transfer Protocol)*: permite transferir arquivos entre dois computadores, em geral, um deles sendo o computador pessoal e o outro um servidor público de arquivos. Para usar esse serviço existe a necessidade de um programa FTP, como, por exemplo, o *WS_FTP*, o *CuteFTP*, o *FTP Explorer* ou o programa "FTP" existente no Windows (executado a partir da linha de comandos). Os servidores de FTP são identificados pelo seu nome de máquina, em geral *ftp.xxx.yy*, ou seja, *ftp* seguido do domínio da instituição que gere o servidor (UNIVERSIDADE..., 2003).
- *Correio Eletrônico ou e-mail*: permite que as pessoas ligadas à Internet troquem mensagens entre si. Atualmente utiliza a Web como suporte. Um endereço de *e-mail* é sempre constituído por duas partes separadas pelo símbolo @, a primeira parte (antes do símbolo) é o nome pelo qual alguém se identifica, a segunda parte (depois do símbolo) identifica qual é o servidor e após o ponto é identificado o local onde o servidor se encontra (UNIVERSIDADE..., 2003).
- *IRC (Internet Relay Chat)*: serviço de mensagem instantânea. Permite uma “conversa” através de mensagens de texto. Para utilizá-lo é necessário um programa. Cada usuário é identificado

por um “nome” (*nickname*), escolhido por ele próprio cada vez que se liga a um servidor de IRC. Esse Serviço pode ser utilizado através da Web (UNIVERSIDADE..., 2003).

Para ilustrar melhor a importância do uso da rede no mundo, ou seja, do estabelecimento definitivo e irreversível dessa tecnologia, a Tabela 1 indica seu número de usuários e o percentual de crescimento ocorrido entre os anos de 2000 e 2003.

Tabela 1. Distribuição mundial de usuários de Internet de 2000 a 2003(em milhões)

	2000	2001	2002	2003	% crescimento		
					2000-2001	2001-2002	2002-2003
África	~ 4.5	~ 6.1	~ 9.9	~ 12.1	34	63	21,38
Ásia	~ 109.2	~ 150.5	~ 211.2	~ 243.4	38	40	15,25
Europa	~ 110.8	~ 143.5	~ 176.2	~ 188.9	30	23	7,24
América Latina e Caribe	~ 17.6	~ 29.2	~ 42.4	~ 44.2	65	45	4,19
América Norte	~ 136.1	~ 156.8	~ 175.1	-	14	12	-
Oceania	~ 8.8	~ 9.6	~ 11.6	~ 11.8	16	21	1,88
Países desenvolvidos	~ 285.4	~ 339.4	~ 388.7	~ 396.7	19	15	2.06
Países em desenvolvimento	~ 94.3	~ 139.3	~ 209.5	~ 246.2	48	50	17,53
Total	~ 387.5	~ 495.8	~ 626.5	~ 675.6	27,96	26,36	7,84

Fonte: Adaptação de Nazareno et al. (2007, p.16).

Já a Tabela 2 apresenta dados relativos ao uso da Internet no mundo em 2005.

Tabela 2. Número relativo de usuários de Internet (2005)

	Percentual de usuários conectados à Internet	% de crescimento de 2000 a 2005
África	2,6	423,9
Ásia	10,4	232,8
Europa	36,4	179,8
Oriente Médio	9,6	454,2
América do Norte	68,6	110,4
América Latina / Caribe	14,7	350,5
Oceania / Austrália	52,6	134,6
Mundo	16,0	189,0

Fonte: Nazareno et al. (2007, p.16).

Podemos afirmar que, um dos serviços responsáveis pelo crescimento de uso da Internet é a Web, juntamente como o correio eletrônico, sendo talvez mais conhecido do que sua infra-estrutura física que acopla alguns dos seus serviços, como o próprio correio eletrônico e o IRC.

4.4 World Wide Web

A proposta inicial da Web é de 1980 e sua primeira concepção ocorreu quando Tim Bernes-Lee escreveu um programa que permite a ligação aleatória de nós (atuais links) (WORLD..., 2005). Certamente, além dessa iniciativa há algumas outras que foram consideradas e serão tratadas mais adiante.

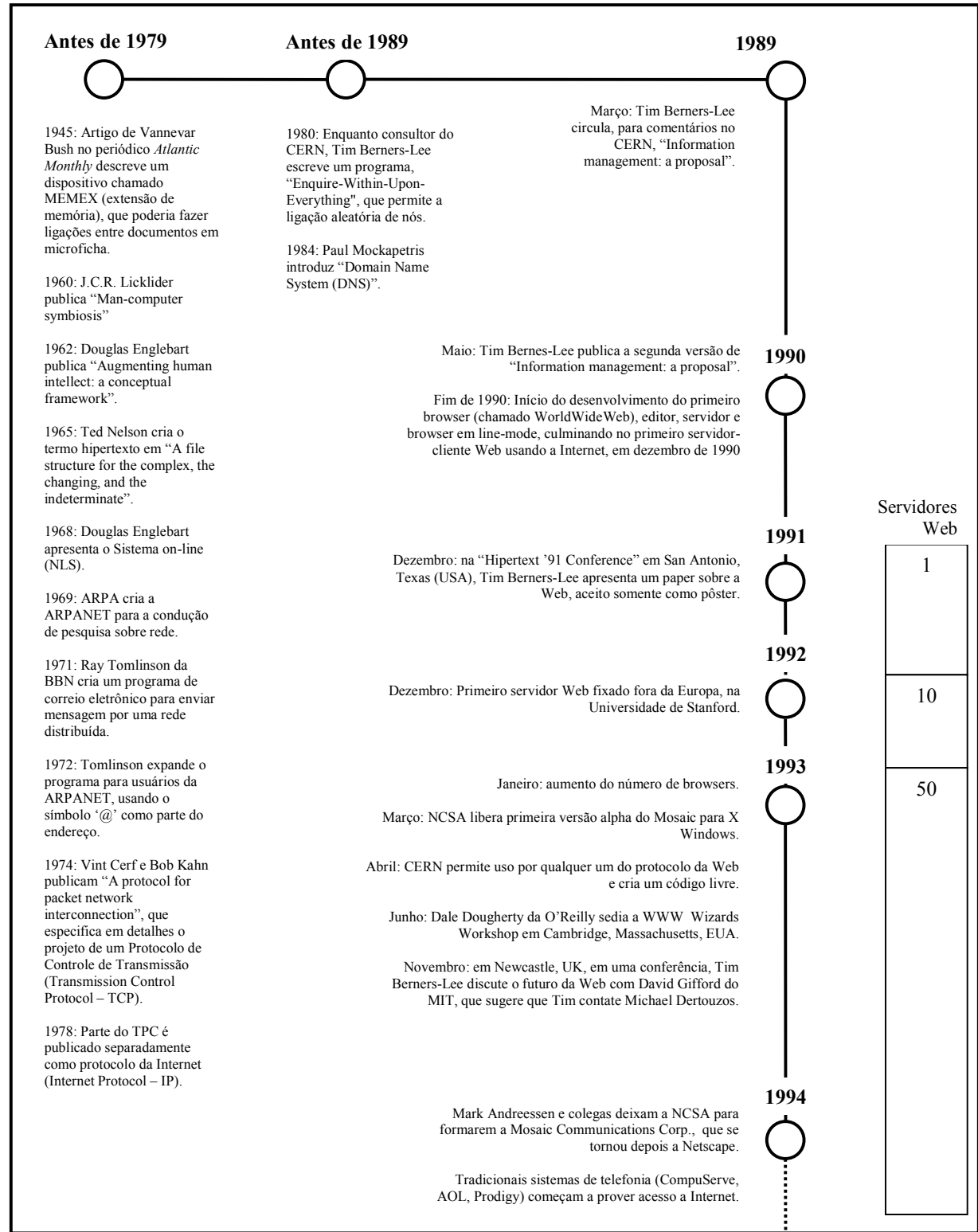
Em 1989 Tim Bernes-Lee publica o artigo “Information management: a proposal” e em 1990, cunha o termo “*World Wide Web*” criando seu primeiro *browser*⁶⁷, editor e servidor (Figura 7).

Em 1994, Berners-Lee cria um consórcio responsável pelo desenvolvimento da Web: o “*World Wide Web Consortium*” e se torna seu diretor. O objetivo principal desse consórcio se baseia no valor social que a Web possui para a comunicação humana, o comércio e as oportunidades de compartilhar o conhecimento.

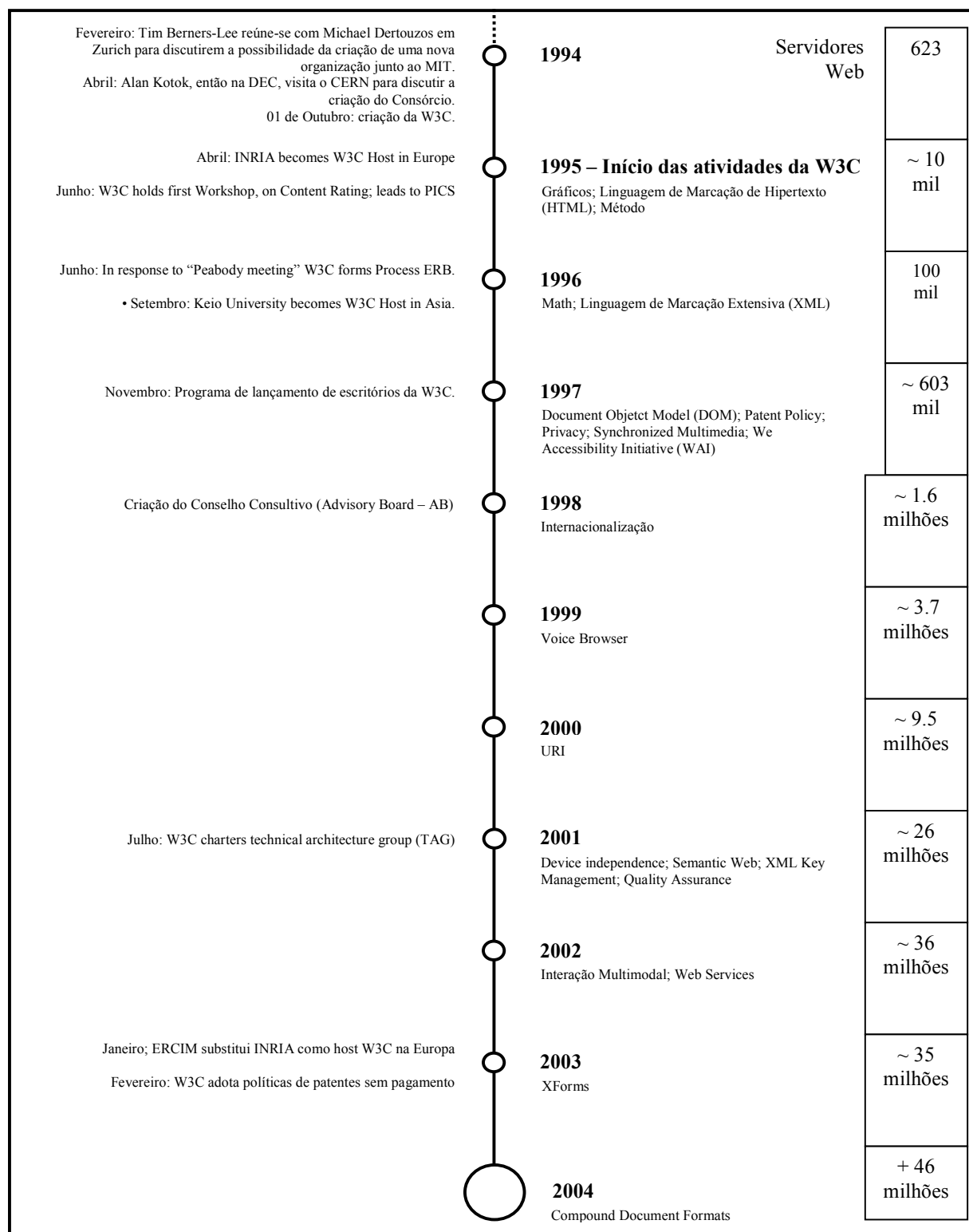
Assim sua meta é “tornar esses benefícios disponíveis a todas as pessoas, independentemente do seu equipamento, software, infra-estrutura de rede, idioma nativo, cultura, localização geográfica ou capacidade física ou mental” (WORLD..., 2008, on-line). As Figuras 10 e 11 mostram toda a trajetória desde do que se considera o marco zero, com a publicação do artigo de Vannevar Bush em 1945, questionando a forma de como a informação é organizada e sugerindo que essa atividade deve ser feita da forma como os seres humanos pensam. A partir daí, apontam-se alguns outros eventos fundamentais para o surgimento da Internet e depois para o da Web, até a criação do W3C e sua trajetória até 2004, incluindo atividades desenvolvidas por essa última, como a WS em 2001.

⁶⁷ Programa utilizado para navegar na Web e que também permite acesso a recursos como o correio eletrônico, transferência de arquivos, etc. (LEÃO, 2001, p.139).

Figura 10. Principais fatos para a criação e o desenvolvimento da Internet e da Web antes da W3C



Fonte: Adaptação (WORLD..., 2005).

Figura 11. Desenvolvimento da Web a partir da criação da W3C em 1994

Fonte: Adaptação (WORLD..., 2005).

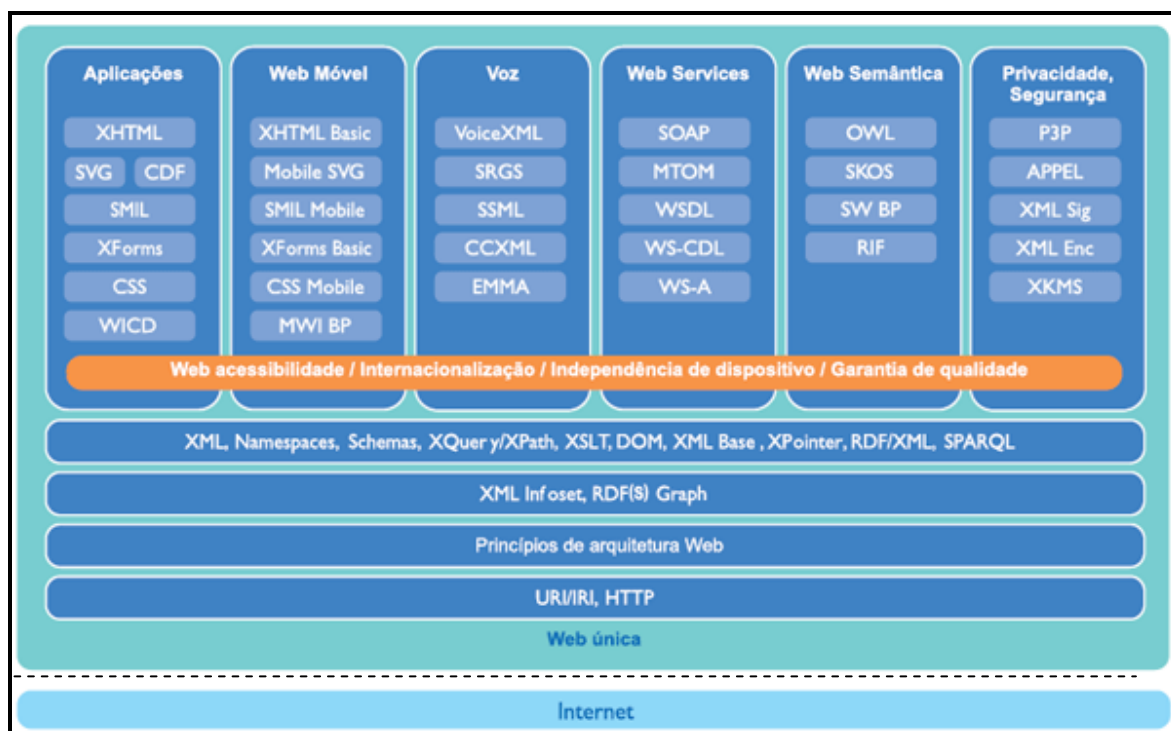
Atualmente, as Atividades do W3C (WORLD..., 2008) para desenvolvimento da Web compreendem:

- Extensible Markup Language (ou Linguagem de Marcação Extensível) (XML): formato de texto simples e flexível derivado do SGML (ISO 8879). A atividade XML procura um equilíbrio, mantendo a estabilidade, fazendo melhorias que estimulam a interoperabilidade.
- Graphics (Imagens gráficas): imagens gráficas desempenham um papel essencial no uso diário da Web, desde desenhos decorativos, passando por publicidade, até gráficos e interfaces gráficas interativas.
- HTML: desenvolvimento dessa linguagem de marcação.
- Internacionalização: seu objetivo é garantir que os formatos e protocolos do W3C estejam abertos a todos os idiomas, sistemas de escrita, códigos de caracteres e convenções locais.
- Math (Matemática): o Grupo de Trabalho para Matemática, que criou e administra a Mathematical Markup Language (MathML – Linguagem de Marcação Matemática), uma codificação em XML estruturada em informações para expressões matemáticas, que facilita a criação e apresentação de expressões matemáticas impressas e na tela, formando uma base para a comunicação matemática de uma máquina para outra no ambiente Web.
- Iniciativa para Web Móvel: hoje o acesso móvel à Web ainda sofre com problemas de interoperabilidade e possibilidades de uso. A Iniciativa para Web Móvel do W3C (W3C MWI) dedica-se a resolução desses problemas.
- Interação Multimodal: procura estender a Web para permitir que os usuários escolham o modo de interação mais adequado às suas necessidades, levando em conta eventuais deficiências, com o objetivo fazer com que os desenvolvedores de aplicativos da Web, ofereçam interfaces eficientes para qualquer modo que o usuário selecionar.
- Política para Patentes: permite que o W3C implante e opere com sucesso sua Política para Patentes, que entrou em vigor em fevereiro de 2004. É importante que a comunidade do W3C esteja organizada para monitorar a aplicação da política e os desdobramentos no ambiente jurídico e normativo.
- Privacidade: desenvolvimento de políticas e tecnologias que permitam a privacidade de usuários comuns e empresas na Web. O centro dessa atividade nos últimos nove anos foi a Plataforma para Preferências de Privacidade (P3P).

- Rich Web Client: reúne o trabalho do W3C sobre Aplicações na Web e formatos de documentos compostos (termo do W3C para um documento que reúne vários formatos, como XHTML, SVG, SMIL e XForms), ou seja, devido a uma grande variedade de plataformas e equipamentos sendo usados, muitos desenvolvedores estão usando a Web como um ambiente de aplicativos. Geralmente esses aplicativos são menores do que aplicativos normais.
- Segurança: essa atividade segue como direções principais a compreensão do usuário quanto aos objetivos e estratégias de outros usuários e aplicativos mal-intencionados e as tecnologias de segurança para XML.
- Web Semântica: sua meta é criar um meio universal para a troca de dados, interconectando a administração de informações pessoais, integrando aplicações em empresas e compartilhando dados comerciais, científicos e culturais em escala global. A partir de instalações para colocar na Web dados que podem ser compreendidos por máquinas, com a ajuda de organizações, indivíduos e comunidades. A Atividade da Web Semântica é uma iniciativa do Consórcio destinada a um papel de liderança na definição da Web.
- Estilo: as folhas de estilo oferecem um amplo controle sobre a apresentação das páginas na Web, um exemplo é a linguagem Cascading Style Sheets (CSS – Folhas de Estilo em Cascata).
- Multimídia Sincronizada: os grupos de trabalho dessa atividade desenvolveram a Linguagem para Integração de Multimídia Sincronizada (Synchronized Multimedia Integration Language – SMIL, como em "smile"), que coordena a combinação apresentações multimídia em que áudio, vídeo, texto e imagens em tempo real. SMIL é uma Recomendação do W3C que permite a especificação e controle do tempo exato em que uma frase é dita, para que coincida com a exibição de uma determinada imagem.
- Aplicativos para Web Ubíqua: atividade lançada em 30 de março de 2007, inclui o Grupo de Trabalho de Geolocalização e o Grupo de Trabalho para Aplicativos da Web Ubíqua. Viabilizando aplicativos que utilizam diversos aparelhos para chegar ao mundo físico, em áreas como segurança, saúde, monitoramento e controle ambiental, entretenimento doméstico, grupos de trabalho distribuídos e manutenção em tempo real.
- Vídeo na Web: objetiva fazer com que os vídeos se tornem efetivamente parte da Web, e não apenas uma extensão que não aproveita totalmente a arquitetura da Web.

- Navegador por voz: objetiva a convergência das telecomunicações e da Web, levando as vantagens da tecnologia da Web para o telefone. O Grupo de Trabalho dedica-se a linguagens para registrar e produzir fala e administrar o diálogo entre o usuário e o computador.
- Departamento do Programa Internacional WAI: o Departamento do Programa Internacional da Iniciativa para Acessibilidade do Conteúdo na Web (WAI) oferece um fórum para que os representantes da indústria, da comunidade de pessoas com necessidades especiais, o setor de pesquisas e o governo trabalhem juntos na identificação de requisitos para acessibilidade e desenvolvimento de soluções seguindo o processo do W3C.
- WAI Técnica: desenvolve as tecnologias a partir das indicações do Departamento do Programa Internacional da Iniciativa para Acessibilidade do Conteúdo na Web (WAI).
- Serviços Web: essa atividade oferece uma forma padrão de interação entre diferentes aplicativos de software, que funcionam em várias plataformas e/ou estruturas, através de características como a interoperabilidade e a extensibilidade, bem como por suas descrições que podem ser processadas em máquinas, por meio do uso de XML.
- XForms: linguagem de marcação que atende às necessidades de formulários eletrônicos, baseando-se no XML e resolvendo problemas de autorabilidade, usabilidade, acessibilidade, independência de dispositivos, internacionalização, integração em diferentes linguagens de *hosts*, e reduz a necessidade de escrita.
- E-Governo: refere-se ao uso da Web ou de outras tecnologias da informação por órgãos do governo na interação com seus cidadãos, entre departamentos e divisões, e entre os próprios governos.

O desenvolvimento dessas atividades acima culmina na estrutura tecnológica da Web, conforme indicado a Figura 8, podendo-se verificar que a Web está sobre a Internet e que todas suas atividades esquematizadas nas colunas refletem seus resultados na Web como um todo.

Figura 12. Estrutura tecnológica da World Wide Web

Fonte: (WORLD..., 2008).

4.4.1 Outros precursores

Além de Bush e Berners-Lee, vale citar outras duas personalidades que, segundo Leão (2001), contribuíram para o que se entende hoje como o ciberespaço⁶⁸ e que foram apresentadas nas figuras 7 e 8: Douglas Engelbart e Ted Nelson.

A partir da premissa apontada por Bush, ou seja, do desenvolvimento de tecnologias que funcionem como apêndice da mente dos seres humanos, Douglas Engelbart cria algumas inovações tecnológicas. Atribui-se a ele o processador de texto, a utilização das redes, a interface de janelas (*Windows*) e o *mouse* (LEÃO, 2001; ROBREDO, 2005). Engelbart fez parte da equipe da ARPA, que colaborou na criação e desenvolvimento da Internet, chamada inicialmente de ARPANET.

⁶⁸ Termo criado por Willian Gibson em 1984, no livro *Neuromancer*. Hoje, é utilizado para designar o espaço criado pelo computador e pelas redes de informação (LEÃO, 2001, p.139; LÉVY, 2000, p.92).

Ted Nelson foi o criador do termo hipertexto, conceito usado por Berners-Lee para a criação da Web. Segundo Leão (2001, p.21) esse termo exprime a possibilidade de manter os pensamentos em sua estrutura multidimensional e não linear. Através do projeto *Xanadu*, Nelson desenvolveu um sistema que pretendia reunir a produção escrita mundial e o compartilhamento de idéias entre pessoas, usando a idéia do hipertexto como base (LEÃO, 2001; ROBREDO, 2005). Robredo (2005) afirma que tal projeto nunca se concretizou.

Outra personalidade importante para a concretização do que é a rede hoje, foi Paul Otlet. Robredo (2005) e Mattelart (2006) fazem menção a ele, afirmando que sua grande contribuição para a idéia da Internet e Web é o espírito universalista. Mattelart (2006, p.45) afirma que “Otlet forja o termo *mundialismo* para melhor marcar a simbiose com um pensamento da rede universal, ao mesmo tempo técnico e social”.

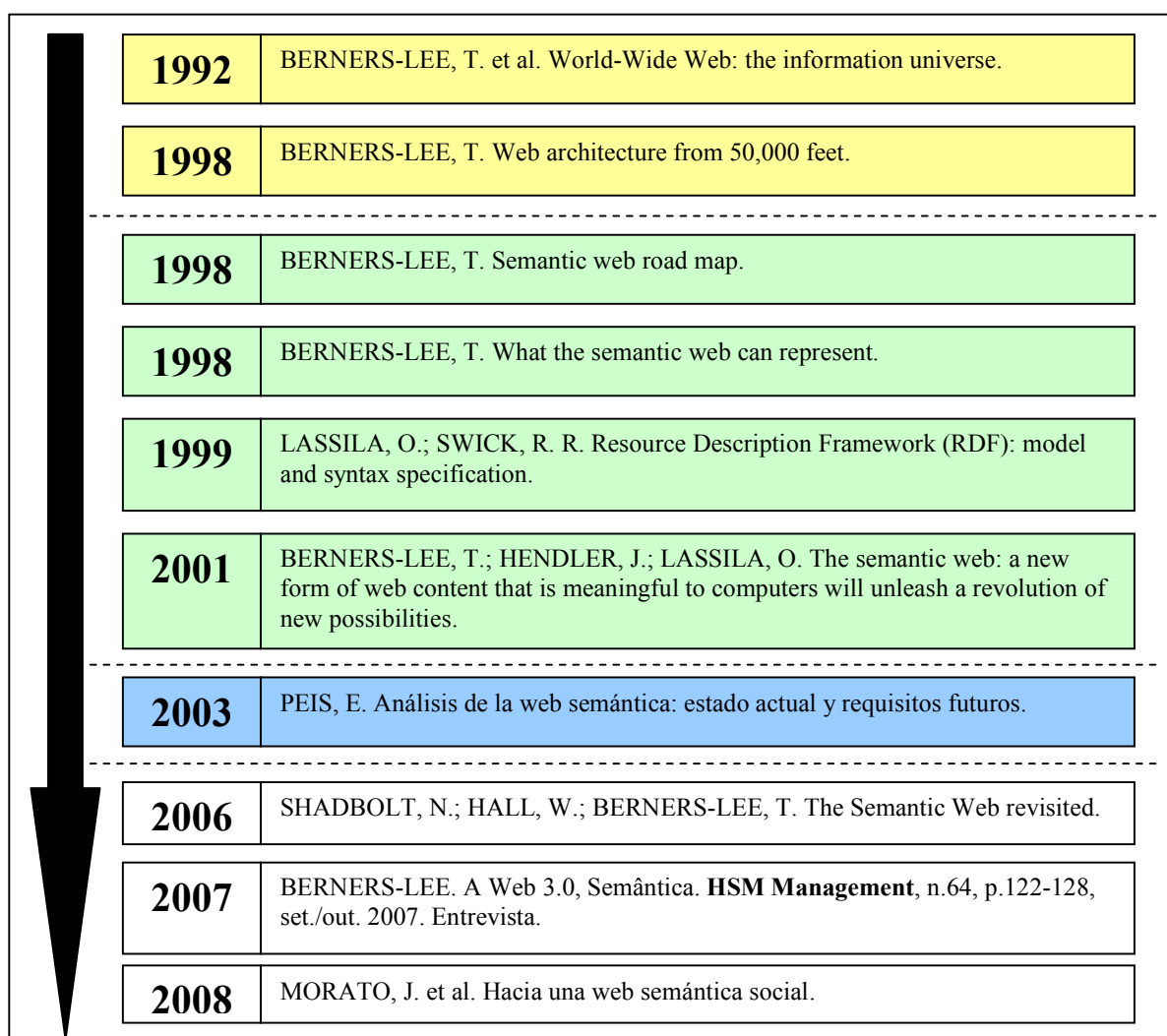
O conceito fundamental da Web é a universalidade, ou seja, ela é concebida com a intenção de que todos possam participar, se quiserem. Segundo Robredo (2001, p.239), a Web é “resultante de várias visões premonitórias e de múltiplos avanços tecnológicos, independente de qualquer projeto específico, e aí residem sua grandiosidade e sua beleza intrínsecas”.

O objetivo desse capítulo, além de apresentar a Internet e a Web de forma breve, foi também de estimular o leitor a refletir sobre o impacto das tecnologias. A CI como uma ciência social aplicada, que tem ligação intrínseca com as TICs, sendo delas originárias, deve deixar por vezes que seu viés humanístico aflore, com o objetivo de contribuir para a elucidação sociedade, mostrando aos indivíduos as diversas possibilidades das propostas relativas à informação e ao conhecimento, mesmo porque, suas relações interdisciplinares também se estabelecem pelo lado sociológico, filosófico, pedagógico, antropológico, histórico, psicológico, etc.

5 PANORAMA DA WEB SEMÂNTICA

Esse capítulo atende ao primeiro objetivo específico: traçar um panorama da WS, através de uma literatura representativa e pertinente sobre o tema, identificando suas propostas e pontos de evolução. A WS é uma extensão da Web atual (BERNERS-LEE, HENDLER E LASSILA, 2001), sendo assim, existe a necessidade de entender como foram as primeiras propostas dessa tecnologia. Para descrever esse objetivo, serão apresentadas as duas primeiras dialéticas indicadas na introdução desse trabalho: 1) Web atual versus WS (movimento, mudança qualitativa); e 2) o movimento inicial até o estado atual da WS (mudança qualitativa). A Figura 13 indica os textos selecionados e como será a apresentação dos movimentos das duas dialéticas expostas acima.

Figura 13. Distribuição dos textos das duas primeiras dialéticas de análise



Os dois primeiros textos, de 1992 e 1998, com preenchimento em cor amarelo claro, oferecem as bases conceituais para o entendimento da Web atual. Eles serão comparados com o segundo conjunto de textos, com preenchimento em cor verde-claro, que entende-se serem as propostas básicas da WS, assim fecha-se a primeira dialética.

Por sua vez, o texto de 2003, com preenchimento em cor azul-claro, pode ser considerado intermediário no entendimento do desenvolvimento da WS. Ele será comparado com o segundo conjunto e com o quarto, que está com preenchimento vazio. Da combinação desses três conjuntos de textos, tem-se a segunda dialética, ou seja, as propostas iniciais da WS até seu panorama atual. Evidente que não se busca aqui estudar o estado da arte da WS, apenas entender o seu desenvolvimento sob o ponto de vista conceitual, como também oferecer bases para a comparação do seu estudo nas teses e dissertações na área da CI no Brasil.

A análise do conteúdo de cada texto individualmente, conforme indicado nos procedimentos metodológicos, com as categorias propostas, encontra-se no Apêndice A. Nesse capítulo, a partir das tabelas no apêndice citado, serão apresentadas de forma direta as duas primeiras dialéticas do trabalho, para evitar redundância dos temas e facilitar a leitura.

5.1 A Web sintática ou atual

A Web inicialmente possuía uma arquitetura e objetivos simples. As idéias que embasam sua criação são: o hipertexto e a recuperação de textos.

Suas principais características são:

- Informações só precisam ser representadas uma vez;
- Estende-se sem “emendas” como notas pessoais na área de trabalho local para bancos de dados grandes em outros locais;
- Os documentos na Web não têm que existir como arquivos: eles podem ser documentos virtuais gerados por um servidor;
- A estrutura em hipertexto permite mais escolhas que os sistemas de menus;
- Sistema altamente descentralizado e tolerante; e

- Simplicidade e mobilidade na engenharia de software.

Pode ser definida como um serviço da Internet, que faz a junção da técnica de hipertexto e recuperação de textos, com o objetivo de facilitar a RI, uma vez que na Web essa recuperação pode acontecer pela dedução de idéias, algo permitido pelo uso dos *links* (nós do hipertexto), por isso Berners-Lee e seus colaboradores consideram o texto de Vannevar Bush como um dos pilares do seu desenvolvimento.

Sua estrutura, portanto, se dá da seguinte forma:

- Índices: documentos por meio do qual se realizam a RI.
- Links: pesquisa de informação por associação de idéias.
- Esquema comum de documentos: uma sintaxe comum que permite que os documentos sejam descritos da mesma forma, facilitando a geração dos índices.
- Protocolo comum de acesso da rede: o HTTP.
- Formato comum de dados para hipertexto.

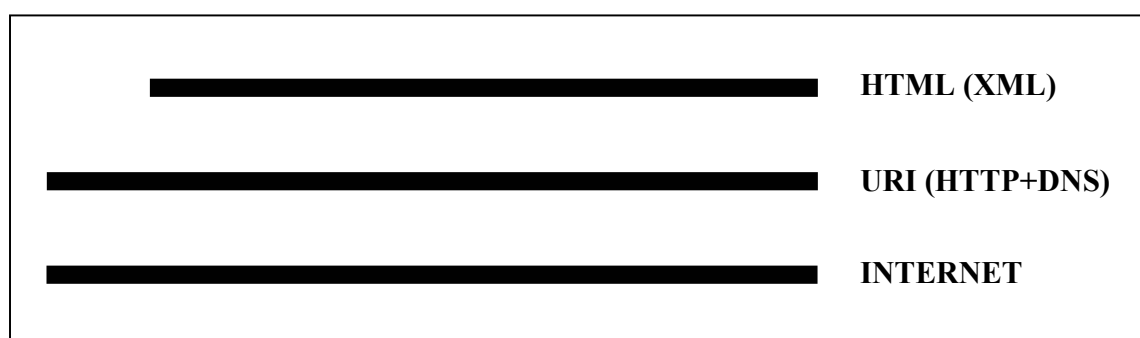
A arquitetura de funcionamento da Web atual deveria ser a seguinte:

- URI (Universal Resource Identifier): identificador universal de recurso, na Web o mais utilizado é a URL (Uniform Resource Locator), conhecido como endereço de Internet.
- HTTP (HyperText Transfer Protocol): protocolo que permite a transferência de páginas na Web, parte da URL.
- DNS (Domain Name System): sistema de nome de domínio, parte do HTTP, única parte da Web centralizada.
- XML (eXtensible Markup Language): linguagem de marcação adotada, mas com uso escasso, devido a facilidade de uso do HTML, que nunca teve local de destaque na arquitetura da Web.
- RDF (Resouce Description Framework): estrutura de descrição de recursos, materializado pela sintaxe XML.

O XML perdeu espaço para uma linguagem mais simples e que se popularizou muito rápido, o HTML. Isso aconteceu pelo modelo ideológico da Web, descentralizado e tolerante, impedindo, por

consequente a uma ampliação do uso do RDF. Não se afirma aqui, que o RDF e o XML estão na arquitetura original da Web, eles aparecem no texto de 1998, ou seja, surgem antes de uma proposta específica e final da WS. Entende-se, portanto que, a arquitetura da Web se estabilizou conforme indica a Figura 14 abaixo.

Figura 14. Arquitetura da Web sintática em 1998



Nesse sentido, um cenário simples de uso da Web nesse estado, se dá quando um usuário consulta um buscador, por meio de um termo (conceitos da recuperação de textos). O buscador consulta o índice em seu banco de dados, devolvendo uma resposta ao usuário, que por sua vez, usa o link para acessar esse recurso e navegar por ele e suas relações, se houverem.

Há, a partir do texto de 1998, mostrado no quadro 13, algumas mudanças de percepção nos objetivos e definições da Web, pois no cenário mostrado, ela apareceu como novidade, mas se mostrou insuficiente em alguns aspectos, como seus objetivos, que sofreram mudanças pra si, ou seja, para sua melhora e desenvolvimento. Pois a RI foi o primeiro e continua sendo o seu principal objetivo, mas como melhorar esse processo, como fazer com que a recuperação seja cada vez melhor.

Assim os objetivos da Web passam a ser:

- A interoperabilidade: qualidade de os sistemas operarem em harmonia, no mesmo padrão.
- Universalidade de acesso, independente de hardware ou plataforma de software, infra-estrutura de rede, idioma, cultura, local geográfico, ou necessidade física ou mental.
- Tornar a informação compreensível por máquinas: tem-se aqui o primeiro rascunho para o projeto da WS.

Pode-se afirmar que a universalidade, a descentralização e a tolerância da Web são suas principais qualidades, mas também o motivo de um certo atraso do seu desenvolvimento, pois o HTML, que nunca teve uma posição definida ou especial em sua arquitetura, toma lugar de destaque entre os usuários, devido à facilidade de uso. Isso é evidenciado, por exemplo, no desenvolvimento dos browsers, pois tiveram que fazer a leitura dos recursos desenvolvidos dessa forma, bem como não faziam a leitura do XML. Assim, pode-se afirmar que todos os indivíduos têm de alguma forma, ligação com o desenvolvimento dessa tecnologia e talvez na WS essa aproximação se estreite ainda mais.

5.2 A Web Semântica

Pode-se afirmar de antemão que a WS é uma resposta a expectativas não alcançadas pela Web no estado em que ela foi concebida. Inicialmente apresentada com uma Web de dados, ou seja, ela seria uma parte da Web em sua primeira versão, logo virou um projeto de enriquecimento da Web atual, isto é, um estágio superior da Web, sua segunda versão (HARMELEN, 2006).

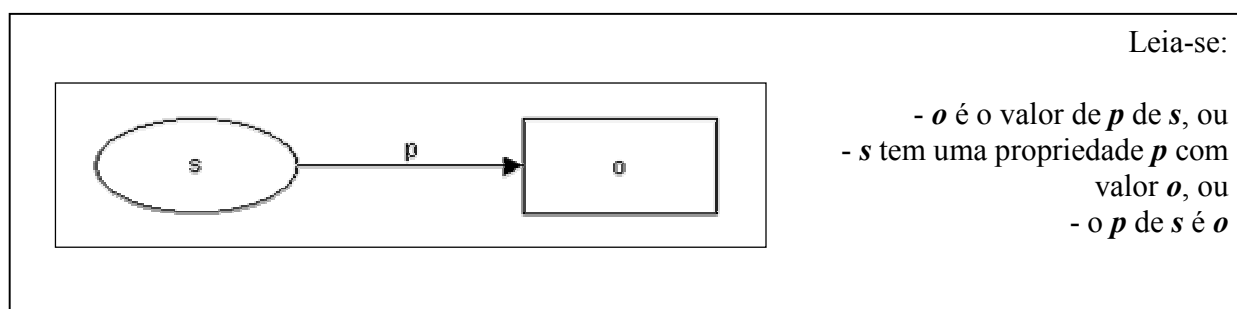
Esboços da WS foram indicados no texto “*Web architecture from 50,000 feet*” de 1998, onde ela parece nessa primeira versão indicada por Harmelen (2006), como uma parte da Web, como umas das atividades de desenvolvimento da Web no todo.

No texto “*Semantic Web road map*” de 1998, é descrito um roteiro para que a WS seja implantada e três indicações são importantes para sustentar sua base:

- As informações na Web são projetadas em sua maior parte para a leitura por seres humanos e não pelas máquinas. Nesse caso, a tecnologia é apenas um intermediário, como uma “televisão”, um canal e não um agente. A WS, portanto, objetiva desenvolver maneiras que permitam que os documentos sejam ‘entendidos pelas máquinas’. Nesse sentido, a tecnologia deixa de ser um canal, ela se torna um ator efetivo.
- Existe a necessidade de um modelo geral de asserções semânticas, que orientem a criação das ferramentas, a descrição nos documentos e suas relações. O ‘*Resource Description Framework*’ (RDF), ou ‘Estrutura de descrição de recursos’ é a resposta para essa necessidade. Trata-se de um modelo de triplas, ou seja, um **recurso** (documento, um vídeo, um áudio, uma página html, etc.)

possui **atributos** (características, propriedades), que possuem **valores**. Segue abaixo um modelo geral do RDF.

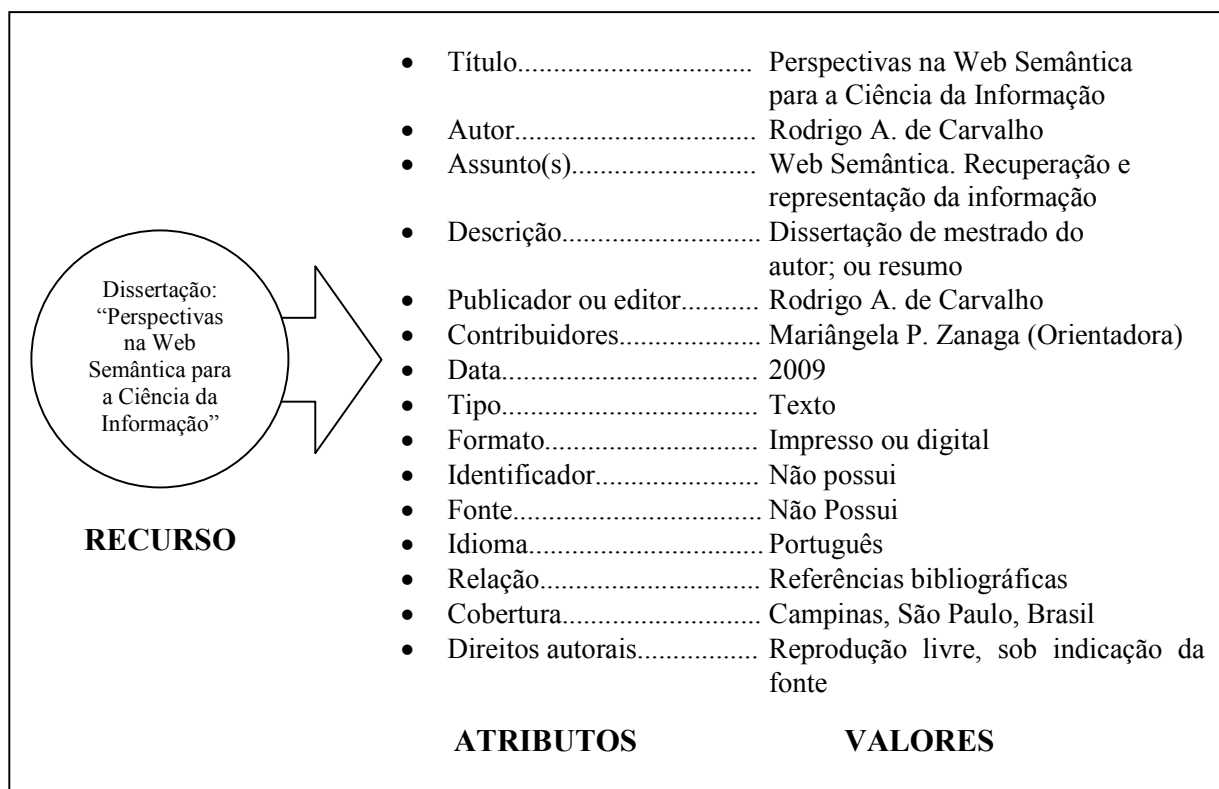
Figura 15. Modelo geral de estrutura de descrição de recursos (RDF)



Fonte: Lassila; Swick (1999).

Um exemplo de aplicação do RDF pode ser feito utilizando o conjunto de metadados DC, que nada mais são do que propriedades gerais para recursos digitais. A figura 16 abaixo ilustra esse exemplo.

Figura 16. Exemplo de aplicação do RDF ao conjunto de Metadados DC



Esse modelo geral de asserção tem vital importância para o desenvolvimento da WS, ela se materializa por meio da sintaxe da linguagem de marcação XML.

- Assinatura digital: os documentos serão analisados não só pelas relações RDF, mas em árvores de asserções sobre quem assinou tais asserções. A validação de prova, por regras de inferência, verificará a lógica, mas pelas asserções pelas quais um documento tenha sido assinado, verificará a assinatura. Esse fator é fundamental para dar legitimidade para as operações na WS, pois determinará se as fontes são confiáveis.

Essas três idéias sustentam o desenvolvimento da WS a partir de então. Mas vale salientar que as principais características da Web na sua origem continuam. Assim, a WS deve se desenvolver de forma universal, tolerante e descentralizada.

A estrutura conceitual da WS, nesse momento, fica da seguinte forma:

- Linguagem de conversão: a aplicação desta camada permite que duas ou mais bases de dados, construídas de forma independente e colocadas na web, estejam ligadas por relações semânticas que possibilitem que as consultas realizadas em uma sejam convertidas em consultas da outra.
- Camada Lógica: necessária para resolver a limitada capacidade dos mecanismos de raciocínio na Web, e definir um subconjunto pleno de RDF que deverá ser compreendido.
- Validação de Provas: os documentos enviados pela Internet serão escritos numa linguagem completa. Contudo, terão de ser restringidos de forma que, no caso de buscas, os resultados sejam computáveis, e na maioria dos casos, sejam provas.
- Linguagens de consultas.

A partir das idéias e camadas conceituais, os objetivos da WS, por meio da necessidade de uma estrutura comum para combinar as informações, nesse momento são: desenvolver um conjunto de aplicações relacionadas a dados na Web; desenvolver linguagens para expressar a informação num formulário processável por máquinas.

Nesse estado da concepção da WS, surgiram dúvidas sobre a ambição e as possibilidades de que elas se realizassem. No texto “What the Semantic Web can represent” de 1998, Berners-Lee, procura esclarecer essas dúvidas, pontuando-as da seguinte forma:

- O conceito ‘documentos compreensíveis por máquinas’, não insinua que a WS objetive fazer com que as máquinas ‘conversem’ com os seres humanos como se fossem ‘amigos’, indica apenas que elas resolverão problemas bem definidos, por meio de operações bem definidas de dados.
- Nesse nível, o RDF se apresenta como uma linguagem (modelo) completa, pois se limita a gerar expressões simples.
- Enquanto bancos de dados relacionais visam tratar com informação de uma especialidade, a WS visa unir dados de muitos modelos diferentes.

Em 2001, Berners-Lee, Hendler e Lassila publicam um esboço final do que é e como deve se desenvolver a WS, a partir do artigo “*The Semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities*”. Nesse texto os autores reforçam o que vinha sendo discutido sobre a WS desde de 1998. Acrescentando algumas justificativas e definindo a sua arquitetura.

Os autores afirmam que a WS vem responder os ‘difamadores’ que afirmavam que a Web nunca seria uma biblioteca bem organizada, sem base de dados centrais e estrutura de árvore, pois jamais haveria a certeza de encontrar tudo. Na estrutura sintática os ‘difamadores’ tinham razão afirma eles.

Algumas condições que são alteradas em relação às anteriores são as seguintes:

- A descentralização será a maior possível, não necessariamente total como na Web sintática.
- Nomeando cada conceito por uma URI, permite que qualquer um expresse a invenção de novos conceitos com mínimo esforço.

Os seus objetivos ficam distribuídos da seguinte forma:

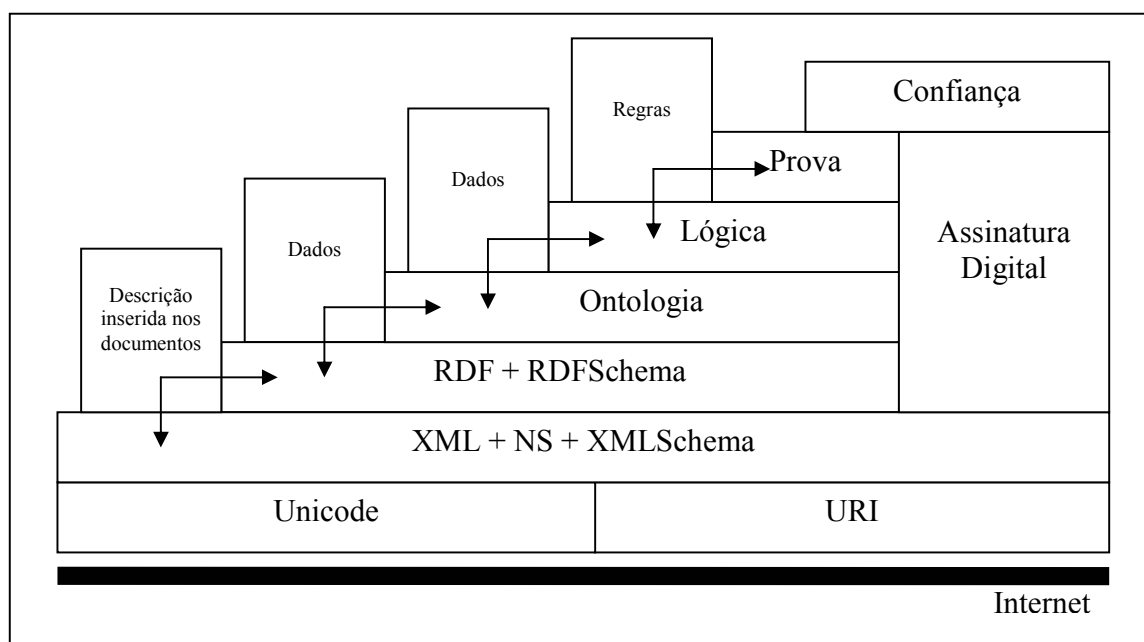
- Oferecer estrutura para o conteúdo significativo das páginas da Web, criando um ambiente em que agentes de software navegando de uma página para outra poderão facilmente realizar tarefas sofisticadas para os usuários.

- Compensar o desequilíbrio no desenvolvimento da Web, que focou mais documentos processados por pessoas do que dados para inferências automáticas.
- Possibilitar um sistema único, porém descentralizado de representação do conhecimento.
- Fornecer uma linguagem que expresse tanto os dados e as regras para raciocínio sobre os dados e que permita que as regras de qualquer sistema de representação do conhecimento existente sejam exportadas para a web.

A definição geral fica da seguinte forma: a WS não é uma Web separada, mas uma extensão da atual, na qual é dado um significado bem definido para a informação, permitindo que computadores e pessoas trabalhem em cooperação.

A arquitetura da WS fica conforme indicado na Figura abaixo.

Figura 17. Arquitetura da Web Semântica



Fonte: Miller (2001).

A partir dessa arquitetura, os softwares chamados agentes inteligentes utilizarão toda a informação semanticamente marcada pelas camadas para solucionar problemas informacionais dos usuários, como também trocar essas informações com outros agentes da Web.

No artigo de 2001, Berners-Lee, Hendler e Lassila apresentam um cenário de aplicação das possibilidades da WS. Nesse cenário, o telefone tocou para um rapaz que ouvia música, imediatamente todos os aparelhos em sua casa que tinham opção de volume abaixaram o som, pois receberam uma mensagem. Sua irmã estava na linha falando de um consultório médico, a respeito da mãe deles, que necessitava consultar um especialista e fazer fisioterapia ao menos duas vezes por semana. A irmã lhe disse que seu ‘agente de web’ estava providenciando encaminhamento. O rapaz se ofereceu para fazer o transporte da mãe. No consultório, a irmã instruía seu ‘agente de web semântica’ através do navegador de seu computador portátil. Rapidamente o agente devolveu as informações sobre o tratamento prescrito à sua mãe a partir do agente do médico, verificando as listas de fisioterapeutas e se eles estavam no plano de saúde da família, num raio de vinte quilômetros da residência da mãe e classificados como excelentes ou muito bom na classificação de serviços confiáveis. Iniciou-se a partir desse momento a tentativa de encontrar uma combinação entre horários disponíveis (fornecido pelos agentes dos fisioterapeutas através de seus sites na Web) e as agendas da mãe e filha. Em pouco tempo seu agente lhe apresentou uma proposta. O rapaz não gostou das opções e colocou seu próprio agente para refazer o plano com preferências mais restritas de localização e horário. O agente da irmã imediatamente forneceu ao agente do rapaz os certificados para acessar os dados que já haviam sido coletados. O plano foi reapresentado e aceito, mas o fisioterapeuta não constava na lista do plano de saúde. O rapaz desligou o telefone e o som voltou ao estado anterior.

Fica evidente a ambição e possibilidades que a WS almeja nesse momento em que muito do que foi divulgado era apenas especulação. No texto “Análisis de la web semántica: estado actual y requisitos futuros” de 2003, Peis et al. fazem essa observação, afirmando que na RI na WS, os agentes inteligentes realizarão tarefas complexas de associação e inferência, devolvendo resultados precisos e contextualizados.

Eles alertam ainda que uma Web aberta e descentralizada pode atrasar o desenvolvimento da WS, que seus desenvolvedores devem buscar o equilíbrio entre facilidade de uso e capacidade de expressão, uma vez que foi a facilidade do HTML que ajudou a difundir a Web.

Nesse texto há um panorama geral da WS e sua estrutura, como também um esboço das características que devem ter os agentes inteligentes na realização de suas atividades:

- Comunicativo: entender as preferências, necessidades e objetivos dos usuários.

- Autônomo: tomar decisões segundo nível de permissão do usuário.
- Adaptável: apreender com os usuários, fontes de informações e outros agentes.

Tem-se assim em um momento intermediário da WS, a discussão sobre agentes e o estabelecimento de algumas linguagens de ontologias. Mas de forma geral suas possibilidades são sempre discutidas, Shadbolt, Hall e Berners-Lee em um artigo de 2006, intitulado “The Semantic Web revisited”, fazem uma revisão de suas propostas e apresentam alguns obstáculos ao seu desenvolvimento.

Segundo os autores muitos duvidam do intento da WS, por não haver produção em larga escala, por ainda não haver mediação baseada em agentes, por não haver qualquer aparência de que a WS esteja presente em qualquer escala. Como a Web sintática transformou o uso do hipertexto em escala mundial, a WS não tem conseguido com o RDF.

Sem normas bem desenvolvidas e acordadas por toda a comunidade a WS não pode existir, além da imprevisibilidade da necessidade de atualização das ontologias e dos acordos terminológicos, dos custos desse empreendimento. Um das soluções apontadas pra isso é a classificação das ontologias em profundas e superficiais, bem como na reafirmação da reutilização de ontologias e dados existentes.

Os autores entendem que muitas perguntas ainda não foram respondidas, indicando-as:

- Como examinar o número elevadíssimo de repositórios de informação descentralizada de escalas variáveis.
- Como alinhar e mapear duas ontologias.
- Como construir um navegador para a WS que efetivamente visualize e navegue no enorme grafo RDF.
- Como estabelecer confiança e procedência dos conteúdos.

Nesse momento da WS, seus objetivos se juntam à possibilidade de construção do que os autores chamam de uma nova ciência da Web e pedem por análises e propostas científicas para sua efetivação como ferramenta de gestão do conhecimento. Berners-lee em entrevista sobre a WS, diz que seu obstáculo está na dispersão dos dados em diversos formatos.

No texto “Hacia una web semántica social” de 2007, os autores fazem uma relação entre a web chamada 2.0 e a WS, e mostram respostas interessantes para um atraso na efetivação da WS. Eles afirmam que o sucesso da Web está na simplicidade das ferramentas, na interação e intervenção dos usuários e na relação com instituições como o W3C, que se comprometem em ser referencia.

A Web 2.0 ou social proporciona as seguintes vantagens: permite que qualquer usuário possa discriminar os termos de descrição do recurso que será disponibilizado e suprime custos da construção de instrumentos terminológicos para um ambiente amplamente heterogêneo como a Web. Essa Web é caracterizada pelos *blogs* e permite essas vantagens de interação para o usuário através do uso das *folksonomias*, que taxonomias pessoais, mas que podem ser compartilhadas. Esse modelo da Web apresenta diversos problemas para a organização da informação e interoperabilidade, como: ambigüidade terminológica; interpretação automática incorreta; dificuldade em expandir ou restringir consultas; dificuldade de incorporação de inferência mediante a representação da informação e de assegurar qualidade na descrição dos recursos.

A WS ou 3.0, segundo os autores, esbarra em obstáculos já superados pela Web social: falta de legibilidade das linguagens semânticas, como RDF; poucas ferramentas amigáveis para a criação de documentos semânticos; e presença de duplicidade nos metadados e ontologias, que provoca desconfiança do usuário.

Pode-se afirmar, que o modelo da Web social no que diz respeito à representação da informação, como um empreendimento comunitário que se estende a todos, é ineficiente e sem grandes perspectivas, mas ela permite a interação dos usuários, e quando eles se envolvem, a tendência é uma implementação garantida de uma tecnologia. Ainda que ele crie ‘vícios’ ruins para efetivar a Web como um sistema de informação.

O quadro abaixo ilustra a relação dessas evoluções da Web.

Quadro 22. Algumas diferenças entre a Web social e a Web Semântica

	Web 2.0	Web semântica
Origem	Evolução natural da Web	Proposta de Berners-Lee para revolucionar a Web.
Implantação	Muito alta	Escassa
Coordenação	Não existe	W3C
Foco	Pessoas	Máquinas
Primeiras menções	2003	1999
Expressão	Linguagem livre, expressado mediante folksonomias, palavras-chave denominadas <i>tags</i> , com problemas de sinonímia e polissemia.	Linguagem controlada, mediante linguagem de expressão de ontologias, sistema de organização do conhecimento e vocabulários de metadados.
Algumas Características	<ul style="list-style-type: none"> - Descrição de recursos para melhorar sua distribuição gratuita - Arquitetura de colaboração - Usabilidade alta - Um recurso é mais útil quanto mais uso tem 	<ul style="list-style-type: none"> - Utilização de uma linguagem padronizada com sintaxe uniforme e semântica não ambígua. - Interoperabilidade: troca de informação entre qualquer repositório. - Usabilidade escassa.

Fonte: Morato, J. et al. (2008, p.81).

O foco para a justificação da Web social é o seu amplo uso, e esse parece um caminho para a WS, se tornar interessante para o público da Web.

Pois bem, a WS tem projeto definido, mas certamente não efetivado, pode-se afirmar que ela está em um ciclo dialético, ou seja, ainda no campo das discussões, das dúvidas, dos conceitos. Existe uma evolução (dialética) clara da Web sintática para a WS, mas esse movimento qualitativo não acontece dentro dela própria, não da forma como se esperava.

Parece necessário que a WS deixe alguns conceitos da Web sintática para poder se desenvolver, ou seja, diminuir a descentralização e setorizar as iniciativas, para no futuro pensar na interligação dos dados.

Quando diz-se de setorizar as iniciativas, seria desenvolver, por exemplo, aplicações semânticas para a informação científica, deixando de tentar atuar, por um tempo, ao alcance da Web como um todo.

6 A WEB SEMÂNTICA NA CIÊNCIA DA INFORMAÇÃO: um estudo a partir das teses e dissertações de programas de pós-graduação da área no Brasil

Esse capítulo atende ao segundo objetivo específico proposto no trabalho, a partir da análise das teses e dissertações defendidas de 2001 até 2008 nos programas de pós-graduação em CI no Brasil, disponíveis na biblioteca de teses e dissertações do Portal da Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) ou nos *sites* de cada programa.

Para efeito de análise, cabe definir o que é uma tese e uma dissertação. A tese é um tipo de trabalho científico monográfico sendo, segundo Severino (2003), o mais representativo. Segundo Marconi e Lakatos (2001, p.166), “apresenta o mais alto nível de pesquisa e requer não só exposição e explicação do material coletado, mas também, e principalmente, análise e interpretação de dados”. Sua principal característica é a exigência de contribuição suficientemente original, ou seja, representar um progresso para a área científica que ela representa (SEVERINO, 2003). O seu principal objetivo como atividade acadêmica é a de obtenção de título de doutor (MARCONI e LAKATOS, 2001).

A dissertação é também um trabalho científico monográfico, mas não tão representativo como a tese. Segundo Marconi e Lakatos (2001), tem caráter didático, pois é o resultado da iniciação à investigação. As principais diferenças em relação à tese, é que a dissertação tem o objetivo da obtenção do título de mestre, o nível de originalidade exigido é menor (SEVERINO, 2001) e a sua elaboração se dá numa fase de formação do pesquisador, anterior ao desenvolvimento da tese.

Estruturalmente são trabalhos parecidos, institucionalizados nos programas de pós-graduação como requisitos parciais para a obtenção dos títulos citados. Geralmente a dissertação se caracteriza pela aproximação e reconhecimento por parte do pesquisador de um problema que lhe interessa, nesse caso a pesquisa é classificada como exploratória (GIL, 2002), por isso o grau de originalidade exigido é menor. Na tese, geralmente, as pesquisas são classificadas como explicativas, pois se trata de um aprofundamento e por vezes de uma intervenção direta do pesquisador diante do que foi anteriormente explorado na dissertação (GIL, 2002).

A relação entre esses dois tipos de monografias científicas não caracteriza que uma seja mais importante do que outra. Diz respeito apenas a fases da formação do pesquisador, que possui

interesses na sua área do conhecimento, materializando os resultados e as questões de suas pesquisas nesses tipos de materiais.

6.1 Dados gerais de produção das teses e dissertações

No Quadro 23 está a distribuição dos programas de pós-graduação em CI no Brasil até dezembro de 2008, assim como a quantidade das teses e dissertações defendidas em cada um deles.

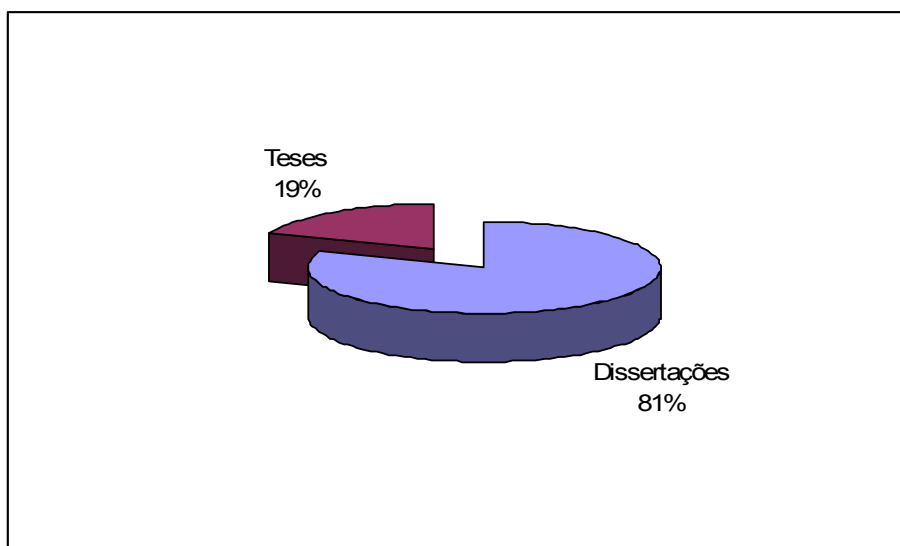
Quadro 23. Distribuição dos programas de pós-graduação em Ciência da Informação no Brasil e das teses e dissertações recuperadas para a pesquisa

INSTITUIÇÃO DE ENSINO SUPERIOR	Dissertações de Mestrado	(%)	Teses de Doutorado	(%)	Totais	(%)
Pontifícia Universidade Católica de Campinas (PUC-Campinas) – SP	01	05,88	Não possui Doutorado	-	01	04,76
Universidade Federal da Bahia (UFBA) – BA	00	00,00	Não possui Doutorado	-	00	00,00
Universidade Federal Fluminense (UFF) – RJ	01	05,88	00	00,00	01	04,76
Universidade Federal de Minas Gerais (UFMG) – MG	01	05,88	01	25,00	02	09,52
Universidade Federal da Paraíba (UFPB) – PB	00	00,00	Não possui Doutorado	-	00	00,00
Universidade Federal de Santa Catarina (UFSC) – SC	04	23,53	00	00,00	04	19,05
Universidade de Brasília (UnB) – DF	04	23,53	02	50,00	06	28,57
Universidade Estadual Paulista (UNESP) – SP	06	35,29	00	00,00	06	28,57
Universidade de São Paulo (USP) ⁶⁹ – SP	00	00,00	01	25,00	01	04,76
Total	17	100,00	04	100,00	21	100,00

⁶⁹ O programa é o de Ciências da Comunicação, com área de concentração em Ciência da Informação, indicado no trabalho recuperado para essa pesquisa.

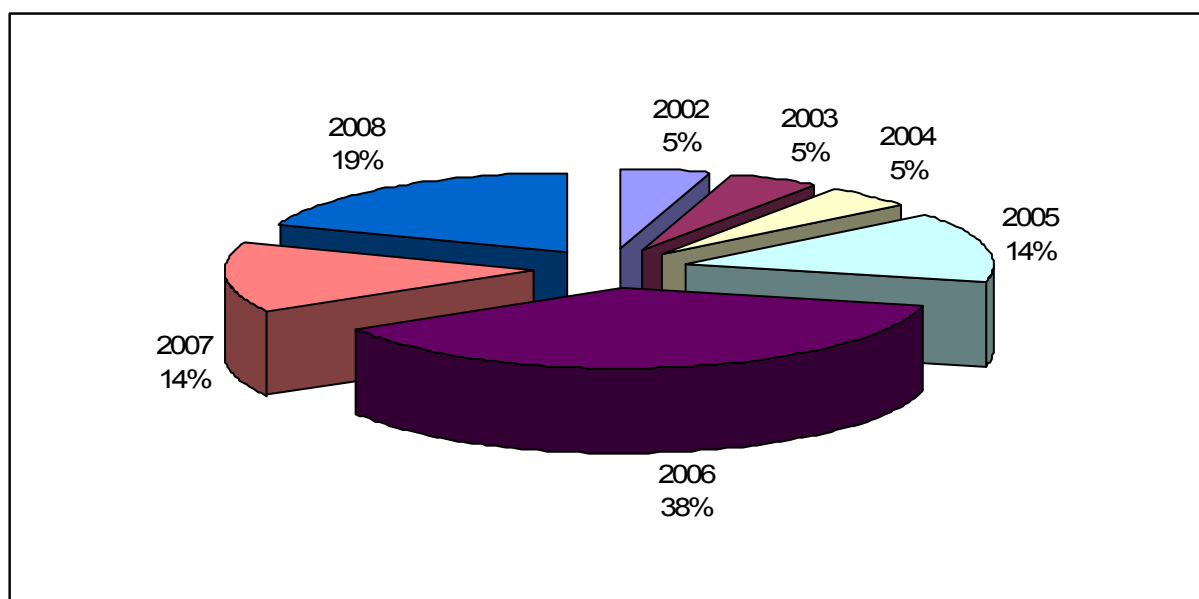
A proporção das teses em relação às dissertações, assim como a distribuição dos anos de defesa de cada trabalho, seguem respectivamente nas Figuras 18 e 19.

Figura 18. Proporção das teses e dissertações recuperadas



Essa diferença na proporção dos trabalhos se deve, principalmente, ao menor número de cursos de doutorado nos programas de pós-graduação.

Figura 19. Distribuição da teses e dissertações por ano de defesa



Na distribuição dos trabalhos pelo ano de defesa não parece haver uma razão para tal resultado.

A identificação das teses e dissertações deu-se a partir de busca pelo termo “Web Semântica”, tendo sido identificados os seguintes trabalhos apresentados nos quadros 24 (teses) e 25 (dissertações) a seguir. Os quadros incluem autoria, título do trabalho, ano de defesa, orientadores e linhas de pesquisa. No apêndice B, seguem as informações sobre as teses e dissertações, incluindo, além dos elementos citados anteriormente, o resumo e as palavras-chave.

O período da pesquisa vai de 2001 até 2008. Para duas dissertações defendidas na UFMG no ano de 2008, não foi possível a consulta ao texto completo, nem aos metadados (resumo, palavras-chave e sumário) para a análise que se propõe, pois não estavam disponíveis on-line, mas apenas citadas em lista no site da universidade, sendo assim, elas não foram colocadas no Quadro 24. Segue a descrição dos trabalhos:

- SILVA, G. B. S. A utilização de mapas de tópicos na compatibilização de conteúdos hipertextuais semanticamente estruturados. 2008. Orientadora: Gercina Â. B. de O. Lima.
- SILVA, D. L. da. Uma proposta metodológica para construção de ontologias: uma perspectiva interdisciplinar entre as Ciências da Informação e da Computação. 2008. Orientador: Renato Rocha Souza.

Quadro 24. Teses da área de CI relacionadas ao tema Web Semântica

UFMG	ALMEIDA, M. B. Um modelo baseado em ontologias para representação da memória organizacional. 2006. Orientador: Ricardo Rodrigues Barbosa. Linha de pesquisa: Gestão da Informação e do conhecimento.
UnB	FEITOSA, A. L. G. A Integração entre sistemas legislativos, terminologia e web semântica na organização e representação da informação legislativa. 2005. Orientador: Haruka Nakayama. Linha de pesquisa: Arquitetura da Informação.
	OLIVEIRA, E. C. Autoria de documentos para a web semântica: um ambiente de produção de conhecimento baseado em ontologias. 2006. Orientador: Mamede Lima-Marques. Linha de pesquisa: não especificada.
USP	FERNEDA, E. Recuperação da informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. Orientador: Johanna Wilhelmina Smit. Linha de pesquisa: Análise documentária.

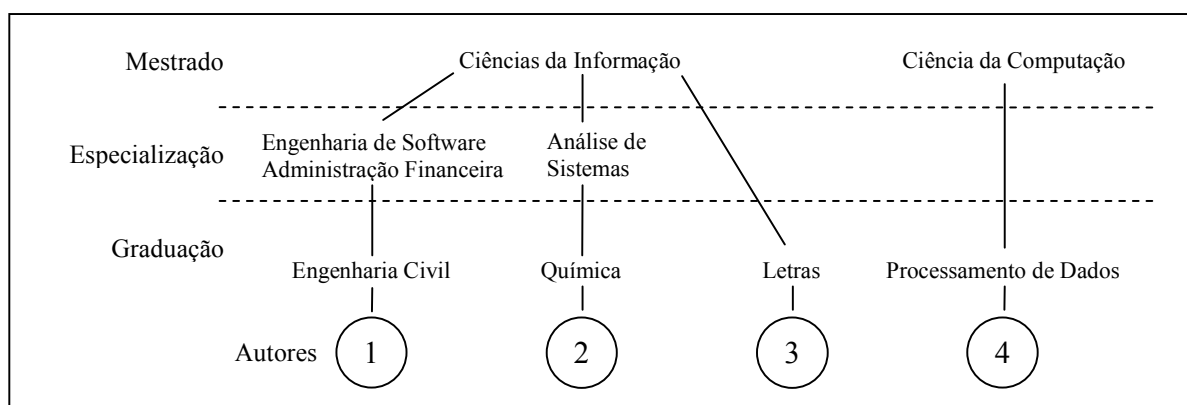
Quadro 25. Dissertações da área de CI relacionadas ao tema Web Semântica

PUC-Campinas	FURGERI, S. Representação de informação e conhecimento : estudo das diferentes abordagens entra a Ciência da Informação e a Ciência da Computação. 2006. - Orientador: Raimundo Nonato Macedo dos Santos; - Linha de pesquisa: Produção e disseminação da informação.
UFF	SALES, L. F. Ontologias de domínio : estudo das relações conceituais e sua aplicação. 2006. - Orientador: Maria Luiza de Almeida Campos; - Co-orientador: Hagar Espanha Gomes; - Linha de pesquisa: Representação, gestão e tecnologia da informação.
UFMG	JORGE, M. A. Ontologias no suporte a portais semânticos . 2005. - Orientador: Marcello Peixoto Bax; - Linha de pesquisa: Gestão da Informação e do conhecimento.
UFSC	SALES, R. de. Tesaurus e ontologias sob a luz da teoria comunicativa da terminologia . 2008. MOLOSSI, S. Inserção da biblioteca digital de teses e dissertações no contexto da web semântica : construção e uso da ontologia. 2008. BEDIN, S. P. M. Metodologia para validação de ontologias : o caso ORBIS_MC. 2007. VITAL, L. P. Recomendações para construção de taxonomia em portais corporativos . 2007. - Orientador: Lígia Maria Arruda Café; - Linha de pesquisa: Fluxos de informação.
UnB	ESBÍZARO, A. L. D. Recuperação de informações sobre log de eventos apoiada em ontologia . 2006. - Orientador: Marisa Bräscher Basílio Medeiros; Linha de pesquisa: Arquitetura da Informação.
	MIRANDA, A. S. S. Ontologias : indexação e recuperação de fotografias baseadas na técnica fotográfica e no conteúdo da imagem. 2007. - Orientador: Mirian Paula Manini; - Co-orientadora: Mamede Lima-Marques; - Linha de pesquisa: Gestão da informação e do conhecimento.
	SANTOS, D. P. L. dos. Um arcabouço teórico para autoria de documentos visando atenuar o surgimento do fenômeno da ambigüidade . 2006. SILVA, L. B. da. Ambigüidades da língua portuguesa : recorte classificatório para a elaboração de um modelo ontológico. 2006. - Orientador: Mamede Lima-Marques; - Linha de pesquisa: Arquitetura da informação.
UNESP	ALVES, R. C. V. Web semântica : uma análise focada no uso de metadados. 2005. CASTRO, F. F. de. Padrões de representação e descrição de recursos informacionais em bibliotecas digitais na perspectiva da Ciência da Informação : uma abordagem do MarcOnt Initiative na era da Web Semântica. 2008. GRACIO, J. C. A. Metadados para a descrição de recursos da Internet : o padrão Dublin Core, aplicações e a questão da interoperabilidade. 2002. - Orientador: Plácida L. V. A. da Costa Santos; - Linha de pesquisa: Informação e tecnologia.
	PINHEIRO, C. B. F. A construção do conhecimento científico : a Web Semântica como objeto de estudo. 2008. - Orientador: Edberto Ferneda; - Linha de pesquisa: Informação e tecnologia.
	RAMALHO, R. A. S. Web Semântica : aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação. 2006. - Orientador: Silvana Aparecida Borsetti Gregorio Vidotti; - Co-orientador: Mariângela Spotti Lopes Fujita; - Linha de pesquisa: Informação e tecnologia; Organização da informação.
	SANTAREM SEGUNDO, J. E. Recursos tecno-metodológicos para a descrição e recuperação de informações na Web . 2004. - Orientador: Silvana Ap. Borsetti Gregorio Vidotti; - Linha de pesquisa: Informação e tecnologia.

6.2 Formação acadêmica dos autores das teses e dissertações

Outras informações que parecem necessárias para entender o estudo da WS na CI, dizem respeito à formação acadêmica de cada autor, pois certamente há uma influência dessa formação nos seus estudos e nas relações interdisciplinares que permeiam a CI. Tais informações foram coletadas nos currículos Lattes dos autores. A Figura 20 apresenta a formação dos autores das teses recuperadas.

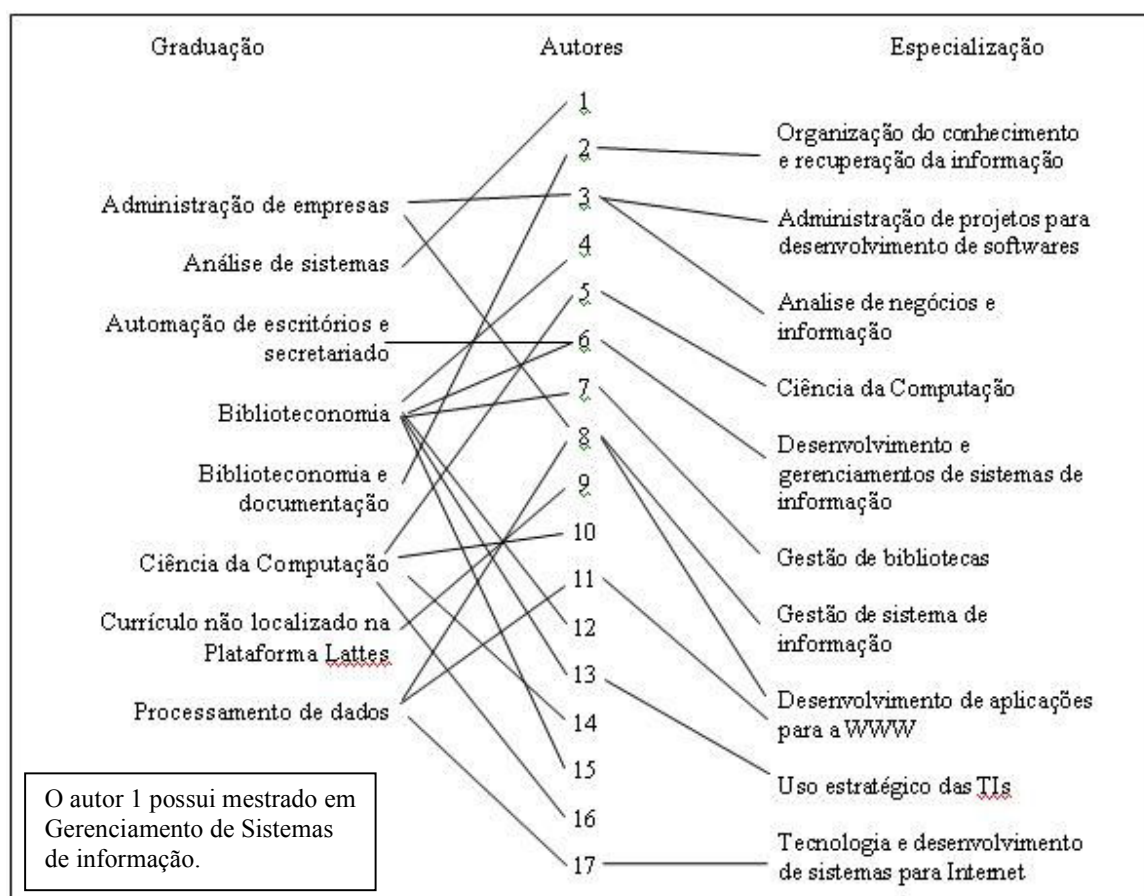
Figura 20. Formação acadêmica dos autores das teses



Percebe-se pela figura, que dos quatro autores das teses analisadas, um tem formação em nível de graduação em Letras e um outro em Processamento de Dados e em nível de mestrado, eles, respectivamente, têm formação em Ciências da Informação e CC, ou seja, estão na relação interdisciplinar proposta pela CI, que pode colaborar para o desenvolvimento da WS.

Os dois primeiros autores possuem formação em nível de graduação que se pode considerar irrelevante para o trabalho, mas possuem especialização em Engenharia de Software e Análise de Sistemas, além de possuírem mestrado em Ciências da Informação, isto é, existe uma ligação interdisciplinar que favorece a justificativa do projeto.

A Figura 21 apresenta a formação dos autores das dissertações.

Figura 21. Formação acadêmica dos autores das dissertações

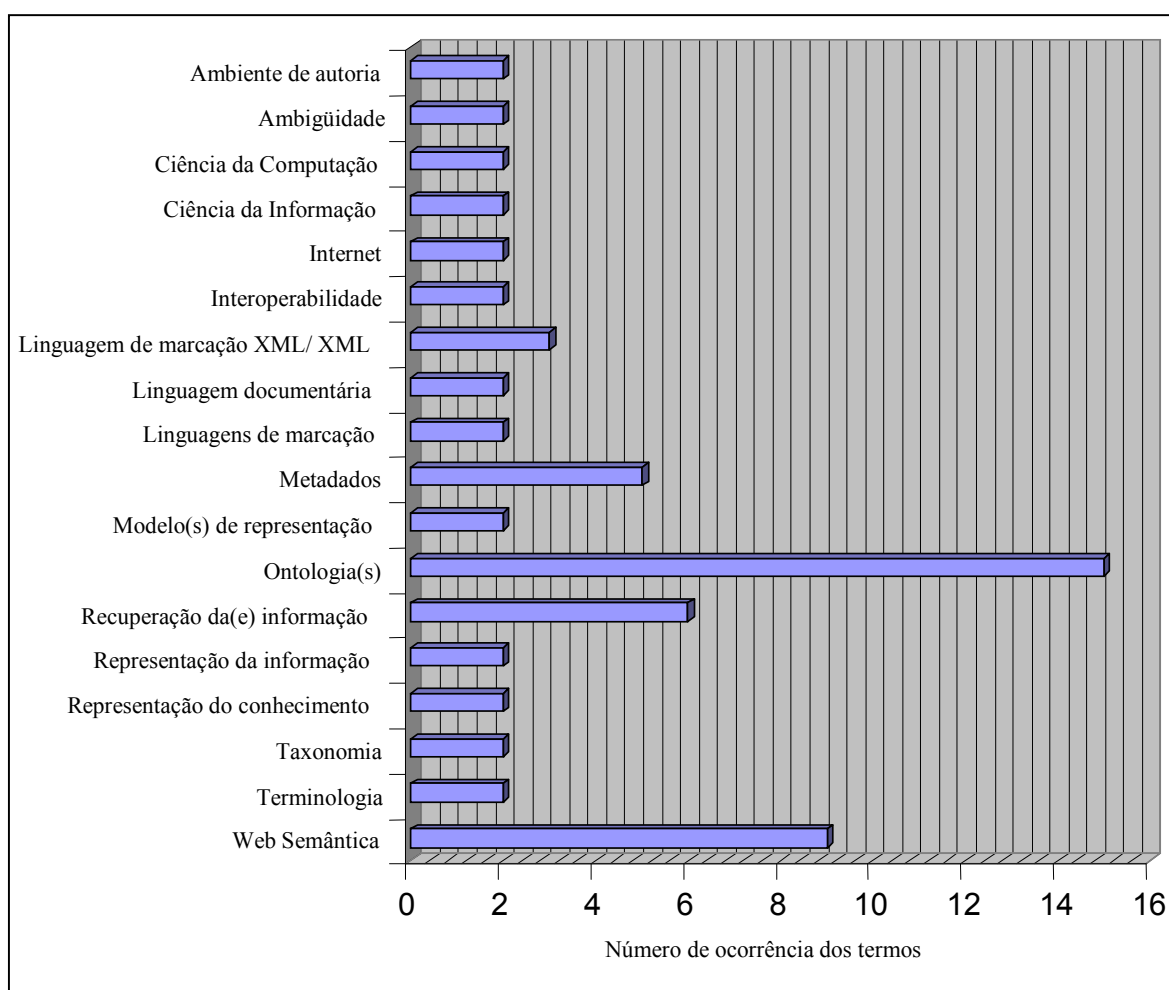
Dos 17 autores das dissertações analisadas, um não teve seu currículo Lattes localizado, 41,18% possuem graduação na área de CI, com bacharelado em Biblioteconomia ou Biblioteconomia e documentação e 47,06% possuem graduação na área de Informática (Análise de Sistemas; CC; e Processamento de Dados).

Em nível de especialização, dos 17 autores, 47,06% não possuem essa formação, 11,76% possuem em áreas da CI (Organização do Conhecimento e Recuperação da Informação; e Gestão de Bibliotecas), 17,65% são especialistas em aplicações e desenvolvimento de tecnologias para Internet e Web (Desenvolvimento de aplicações para Web; e Tecnologia e desenvolvimento de sistemas para Internet), um tem especialização em Ciência da Computação e 17,65% tem formação nesse nível em TICs e Sistemas de Informação (Desenvolvimento e gerenciamentos de sistemas de informação; Gestão de Sistemas de Informação; e Uso estratégico das TI).

6.3 Distribuição das palavras-chave

A Figura 22 apresenta as palavras-chave encontradas nos 21 trabalhos consultados, com ocorrência igual ou superior a dois, sendo que no apêndice C segue a lista completa de todas elas.

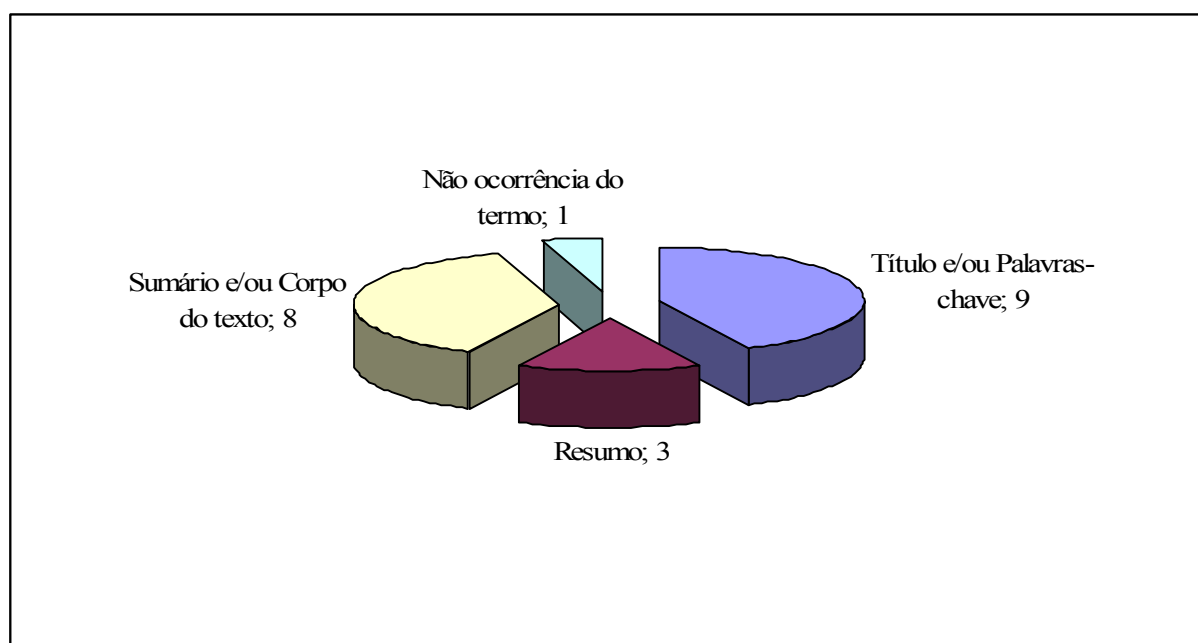
Figura 22. Distribuição das palavras-chave das teses e dissertações



O termo com maior ocorrência é ‘Ontologia (s)’, seguido por ‘Web Semântica’, ‘Recuperação da Informação’, ‘Metadados’ e ‘XML’. Essa tendência reforça a relação da CI com a WS, pela RI e pela representação da informação (metadados e XML) e representação do conhecimento (ontologia(s)).

Para complementar a análise, foi identificada a parte do trabalho (título, resumo, palavras-chave, texto e sumário) do trabalho em que apareceu o termo ‘Web Semântica’, haja vista, que esse comportamento pode indicar a importância do tema no estudo.

Figura 23. Identificação de parte do trabalho onde ocorre o aparecimento do Termo ‘Web Semântica’



Pela Figura 23 acima, verifica-se que 42,86% dos 21 trabalhos contem o termo principal da presente pesquisa no título e/ou palavras-chave, o que indica uma maior proximidade com a WS. A ocorrência do termo no resumo dos trabalhos ocorre em 14,29% e no sumário e/ou corpo do texto ocorre em 38,10%.

Em um dos trabalhos, o termo ‘Web Semântica’ não está presente, em nenhum dos pontos de acesso citados, nem mesmo no corpo do texto. O mesmo aparece nas análises, pois foi recuperado na Biblioteca de Teses e Dissertações do Portal CAPES pela expressão exata ‘Web Semântica’. Ou seja, qualquer usuário da biblioteca que realizar a pesquisa pelo termo ‘Web Semântica’ encontrará esse trabalho.

Outra relação de interesse para essa pesquisa, diz respeito à identificação das palavras-chave que ocorrem juntamente com o termo ‘Web Semântica’, nos nove trabalhos indicados acima que trazem o termo no título e/ou palavras-chave. O Quadro 26 abaixo indicando tal relação.

Quadro 26. Palavras-chaves dos trabalhos que se relacionam com o termo ‘Web Semântica’

Ambiente de autoria Ambigüidade Arquitetura da informação Arquitetura de metadados RDF (2) BDTD Bibliotecas digitais Catalogação automatizada Cienciometria Comunidade científica Descrição e recuperação de informação na web Ferramentas de busca Gestão de Recursos Informativos IBICT Indexação Informação e tecnologia Informação legislativa Internet (2) Linguagem de marcação XML (3)	WEB SEMÂNTICA	Linguagem documentária Linguagens de marcação Linguagens de ontologias MARC21 MarcOnt Metadados (2) Ontologia (s) (6) OWL Padrões de metadados Produção de documento Recuperação da(e) informação (2) Representação da informação (2) Representação do conhecimento Sistema legislativo Taxonomia TEDE Terminologia (2) Web
---	------------------	---

Os destaques apontam, novamente, que a relação da WS com a CI acontece pela RI, pela representação da informação e do conhecimento.

Outros conjuntos de palavras-chave parecem se voltar mais especificamente para a área de graduação de seus autores. Através delas, procura-se verificar a influência da área de formação nos estudos realizados em nível de pós-graduação. O Quadro 27 abaixo apresenta a distribuição das palavras-chave dos autores de dissertações com formação na área de Biblioteconomia.

Quadro 27. Distribuição das palavras-chave das dissertações com autores com formação em nível de graduação na área de Biblioteconomia

Arquitetura de metadados RDF Bibliotecas digitais Catalogação automatizada Cienciometria Comunidade científica Informação e tecnologia Linguagem de marcação XML Linguagem documentária MARC21	MarcOnt Metadados Método de análise de conteúdo Modelo de representação Ontologia(s) (4) Organização da informação Padrões de metadados Portal corporativo Recuperação da(e) informação (2)	Relações conceituais Representação da informação (2) Representação do conhecimento Taxonomia Teoria comunicativa da terminologia Tesauro Validação de ontologias Web Web Semântica (3)
--	---	---

Os termos em destaque mostram uma tendência na CI de realizar estudos de recuperação e representação da informação, como também de representação do conhecimento (ontologia(s)).

Vale salientar que dos sete trabalhos que contém os descritores do quadro acima, apenas três possuem o termo ‘Web Semântica’ no título ou nas palavras-chave. Nos outros quatro, o termo aparece apenas no corpo do texto. Tal observação é necessária, pois como foi visto na seção sobre indexação, quando um termo aparece em um dessas áreas, geralmente ele tem uma importância primária no conteúdo do trabalho.

O Quadro 28 abaixo apresenta as palavras-chave dos autores tanto de teses como de dissertações, com formação em nível de graduação na área de informática (Análise de Sistemas, CC e Processamento de Dados).

Quadro 28. Distribuição das palavras-chave dos trabalhos de autores com formação em nível de graduação na área de informática

Ambiente de autoria Ambigüidade (2) BDTD Busca e recuperação de informação na Web Ciência da Computação (2) Ciência da Informação (2) Descrição de recursos na Internet Descrição e recuperação de informação na web Dublin Core Ferramentas de busca Gestão de Recursos Informacionais Gramática de valências IBICT Informação Internet (2)	Interoperabilidade Linguagens de marcação (2) Linguagens de ontologias Log de eventos Metadados (3) Modelos de recuperação de informação Modelos de representação Ontologia (6) Recuperação da(e) informação (4) Taxonomia TEDE Terminologia Web Semântica (4) XML
--	---

Dos trabalhos que contém os termos que compõem o Quadro acima, em quatro o termo ‘Web Semântica’ aparece no título ou nas palavras-chave e em um, o termo aparece no resumo. O comportamento dos termos em destaque sugere uma similaridade com os trabalhos dos graduados na área de Biblioteconomia.

Segue, uma análise simples e individual de cada trabalho onde o termo ‘Web Semântica’ ocorre no título e/ou palavras-chaves, bem como de cada trabalho com a ocorrência do termo no resumo. Entende-se que quando um termo ocorre em algumas dessas partes, geralmente ele tem uma importância mais significativa para o trabalho.

A descrição de cada análise identifica os objetivos de cada trabalho e busca entender como a WS se apresenta no contexto.

6.4 Análise individual dos trabalhos com a ocorrência do termo ‘Web Semântica’ no resumo

O título do primeiro trabalho analisado é “Representação de informação e conhecimento: estudo das diferentes abordagens entra a Ciência da Informação e a Ciência da Computação”, dissertação de mestrado defendida em 2006, que tem como objetivo principal “pesquisar estruturas de representação para recursos informacionais que possibilitem a recuperação de informação e conhecimento de maneira mais efetiva”, numa abordagem entre a CI e a CC. Os objetivos específicos são “estudar os elementos bibliográficos fundamentais necessários à representação da informação e do conhecimento do ponto de vista da CI e da CC”; “pesquisar a existência de pontos convergentes e divergentes da CI e da CC com relação à representação do conhecimento”; e “realizar uma revisão da literatura referente à representação do conhecimento nas duas áreas citadas”. O autor apresenta a WS como uma Web que permite a relação interdisciplinar entre a CI e CC e que exigirá do profissional da informação uma atividade mais direta com o desenvolvimento da Web, dando ênfase para a terminologia e construção de ontologias.

O segundo trabalho analisado é “Ontologias no suporte a portais semânticos”, dissertação de mestrado defendida em 2005. Tem como objetivo principal é investigar e compreender na prática, “se o uso de uma ontologia, cuja construção também é proposta, apóia a organização da informação em portais (hipótese da pesquisa)”; como objetivos específicos: “determinar as principais vantagens e dificuldades do uso da ontologia em portais”; “levantar e analisar as principais ferramentas baseadas em ontologias apresentadas na literatura científicas sobre portais Web”; e “propôs-se à arquitetura funcional de um portal educacional para uma comunidade universitária”. O autor propõe

um estudo de caso, seu foco é a construção de uma ontologia, assim, ele utiliza os padrões propostos pela WS como base da sua revisão de literatura, descrevendo todas as suas linguagens de construção.

O terceiro e último trabalho com ocorrência do termo ‘Web Semântica’ no resumo é “Ontologias: indexação e recuperação de fotografias baseadas na técnica fotográfica e no conteúdo da imagem”, dissertação de mestrado defendida em 2007. O objetivo principal do trabalho é a propor de um modelo para descrição de imagens com base em ontologias, para a melhora da recuperação de imagens em sistemas de busca. Os objetivos específicos são: “propor uma ontologia de descrição de imagens” e um “um sistema protótipo de descrição e busca de imagens com base na ontologia”. O autor utiliza os padrões da WS, ou seja, ele usa o cenário que essa nova Web propõe para a representação de recursos digitais, a partir das linguagens *OWL* e *RDF Schema*. Entre suas conclusões, parece válido salientar, que ele entende que o padrão de metadados DC, não cobre as necessidades de representação considerada adequada para imagens, assim a partir de algumas técnicas de representação de imagens por conceitos, elabora-se uma ontologia.

6.5 Análise individual dos trabalhos com a ocorrência do termo ‘Web Semântica’ no título ou nas palavras-chave

Dos trabalhos em que o termo WS aparece no título ou na palavra-chave, está o intitulado “A Integração entre Sistemas Legislativos, Terminologia e Web Semântica na organização e representação da informação legislativa”, tese de doutorado defendida em 2005. Esse trabalho objetiva identificar como a Terminologia e a WS podem contribuir para a indexação e RI em Sistemas de informação Legislativa. Como objetivos específicos estão: identificar e sistematizar os fundamentos metodológicos e teóricos, referentes a conceito, a sistemas de conceitos, a controle terminológico e conceitual, a linguagem documentária e a WS, para aplicação na organização e na RI nos sistemas de informação legislativa; identificar e sistematizar as características técnicas dos sistemas de informação legislativa; identificar e sistematizar os atributos de controle terminológico e conceitual na organização e na RI nesses sistemas; e elaborar proposta metodológica para a organização, a representação e a recuperação de normas legislativas que: a) utilize controle

terminológico e conceitual na organização e na recuperação das informações; b) utilize tecnologias da WS. A WS aparece nesse trabalho como indicado na última parte da oração anterior, a partir das propostas de suas tecnologias na organização da informação na Web, focada principalmente no que diz respeito às ontologias, que permite a comunicação não apenas entre as pessoas, como a Terminologia e Linguagem Documentária, mas também entre máquinas, facilitando na construção dos sistemas de informação que utilizam a Web como ambiente.

Na tese de doutorado defendido em 2006, intitulada “Autoria de documentos para a Web Semântica: um ambiente de produção de conhecimento baseado em ontologias”, o autor tem como objetivo principal a criação de uma especificação de alto nível, a partir de um novo modelo de arquitetura de sistemas de informação para ambientes de autoria de documentos, baseados em ontologias para a WS. Os específicos são: i) levantar critérios e modelos de análise de sistemas de autoria, de avaliação de ontologias, de conformidade com o W3C e de análise e avaliação de software; ii) realizar uma análise de requisitos de sistemas de autoria baseados em ontologia com base em critérios levantados; e iii) especificar requisitos e modelar a arquitetura de um novo ambiente de autoria baseado em ontologias para a Web Semântica. A WS nesse trabalho é vista como uma área de atuação da pesquisa na CI, ou seja, o autor propõe-se uma intervenção direta no desenvolvimento da WS. Isso é evidenciado pela supervisão do Prof. Dr. Frank van Harmelen no desenvolvimento da tese, bem como por trabalho sobre o tema, apresentado no ‘*International Semantic Web Conference*’ (ISWC) em 2004, como o título de ‘*A framework for ontology-based authoring environments*’. Onde foi definido pela primeira vez o que é o ambiente de autoria:

Um conjunto de ferramentas de escrita, edição e representação de documentos que interativamente apóiam os usuários (autores), permitindo um melhor acesso, uso e representação semântica do conhecimento durante a autoria, por meio das seguintes tarefas: fazer anotação semântica de documentos; criar metadados; ligar os termos do documento com as ontologias externas; ligar documentos similares entre si; transformar citações em *links* rotulados; tornar explícita a estruturação retórica de um documento; e ligar documentos a comentários de revisão de pares (OLIVEIRA; HARMELEN; LIMA-MARQUES, 2004⁷⁰ citados por OLIVEIRA, 2006, p.125).

⁷⁰ OLIVEIRA, E. C.; HARMELEN, F. v.; LIMA-MARQUES, M. A framework for ontology-based authoring environments. In: ISWC 2004 – International Semantic Web Conference, Hiroshima, Japão. 2004.

Na dissertação de mestrado “Inserção da biblioteca digital de teses e dissertações no contexto da Web Semântica: construção e uso da ontologia” defendida em 2008, o autor tem como objetivo principal analisar as tecnologias de RI à luz da WS, para o aperfeiçoamento do método de RI adotado pela Biblioteca Digital de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação Científica e Tecnológica, a partir dos seguintes objetivos específicos: caracterizar no contexto da RI, a proposta da WS; caracterizar a ontologia como ferramenta da WS, destacando suas características, classificações, vantagens, desvantagens e principalmente alguns critérios que devem ser utilizados no seu desenvolvimento; verificar a estrutura atual da BDTD e métodos para RI; elaborar parte da ontologia para o domínio correspondente à intersecção entre a RI e a WS para a BDTD; sugerir a integração do modelo proposto pela WS ao cenário atual da BDTD; propor recomendações na RI da BDTD, baseadas nas tecnologias da WS. O foco do estudo foi a construção de uma ontologia, instrumento fundamental para a WS, essa aparece novamente como uma tendência de padrões, para a construção de sistemas de informação no ambiente com mais semântica, para facilitar a RI.

Em “Ambigüidades da língua portuguesa: recorte classificatório para a elaboração de um modelo ontológico”, dissertação de mestrado defendida em 2006, o autor tem como objetivo geral propor uma classificação dos tipos de ambigüidades presentes na língua portuguesa, no contexto brasileiro, representá-la com uso de ontologias, para gerar uma ontologia de ambigüidades”, a partir dos seguintes objetivos específicos: identificação os tipos de ambigüidades da língua portuguesa no contexto brasileiro; uniformização da interpretação das ambigüidades por meio de um tratamento terminológico; proposição de uma classificação dos tipos de ambigüidades da língua portuguesa, no contexto brasileiro; e a representação da classificação das ambigüidades pela linguagem ontológica da Web – OWL. A WS nesse trabalho aparece como uma tendência para a criação de padrões de organização da informação em ambiente Web. O foco do trabalho é a criação de uma ontologia.

Na dissertação de mestrado “Web Semântica: uma análise focada no uso de metadados”, defendida em 2005, o autor objetiva de forma geral estudar os aspectos da construção da WS em relação com a CI, o que diz respeito ao uso e aplicação de metadados para a representação de recursos informacionais, a partir dos seguintes objetivos específicos: analisar as ferramentas tecnológicas que compõe a WS, destacando para o uso das ontologias para garantir a semântica dos dados; identificar

as características da linguagem de marcação XML – *eXtensible Markup Language*, e sua importância para a estruturação dos recursos informacionais na WS; analisar o uso da arquitetura de metadados, em especial a arquitetura RDF – *Resource Description Framework*, para o estabelecimento da interoperabilidade sintática, estrutural e semântica; e analisar as características, o uso e a aplicação dos diferentes tipos de metadados e seus formatos para o estabelecimento da representação de recursos informacionais. A WS aparece como objeto principal desse trabalho, na busca de convergências no que diz respeito ao uso de metadados para a representação da informação na Web, mas também apresenta um panorama de ligação da CI com a WS.

O trabalho “Padrões de representação e descrição de recursos informacionais em bibliotecas digitais na perspectiva da Ciência da Informação: uma abordagem do MarcOnt Initiative na era da Web Semântica”, dissertação de mestrado defendida em 2008, tem como objetivo geral identificar e estudar métodos e ferramentas tecnológicas disponíveis para a representação e a descrição de recursos informacionais no contexto das bibliotecas digitais e da Web, na tentativa de identificar o embasamento conceitual subjacente às ferramentas para a padronização da descrição dos aspectos de forma e do conteúdo, na abordagem da CI. Os objetivos específicos são: a identificação e caracterização do MarcOnt Initiative, como um padrão de representação da informação (forma e conteúdo) em bibliotecas digitais e na Web na atualidade; justificar a adoção da ferramenta MarcOnt como possibilidade de proporcionar a interoperabilidade semântica entre bibliotecas digitais e outros sistemas de informação. A WS aparece como a tendência futura para o tratamento da informação no ambiente Web, com possibilidades de atuação da CI, e sustenta teoricamente também a ferramenta MarcOnt.

Em “A construção do conhecimento científico: a Web Semântica como objeto de estudo”, dissertação de mestrado, defendida em 2008, o autor objetiva de forma geral refletir sobre a participação e integração da CI e da CC na construção e desenvolvimento do conhecimento científico sobre a WS, a partir da: identificação da produção científica em WS disponível nas principais iniciativas brasileiras de bancos e bases de dados científicos em relação as matrizes disciplinares, canais de comunicação, colaboração científica; identificação dos autores mais citados nas dissertações e teses sobre o tema, para o mapeamento das influências recebidas no seu desenvolvimento científico; identificação da presença da CI nos canais de comunicação da CC e

vice-versa; e a identificação da colaboração científica da CC e da CI através de co-autorias. A WS é o objeto principal desse estudo, no que diz respeito ao seu estudo nos campos disciplinares da CI e CC. Como consideração interessante está o isolamento da CI na produção do conhecimento sobre tema, no que se refere às teses e dissertações, ou seja, a CC, principal disciplina no desenvolvimento da WS, desconhece o material da CI.

Na dissertação de mestrado “Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação”, defendida em 2006, o autor objetiva de forma geral contribuir para uma “desmistificação” dos conceitos e tecnologias subjacentes ao projeto WS, identificando pontos de convergência entre as novas abordagens computacionais e os métodos de análise, síntese, organização e representação de recursos informacionais utilizados tradicionalmente no âmbito da CI e indicando possibilidades de desenvolvimento de novos métodos de organização e recuperação de informações, que favoreçam a junção teórica da CI com as novas tecnologias emergentes da área da CC. Assim, a WS aparece no trabalho como ponto de convergência dessas duas ciências.

Em “Recursos tecno-metodológicos para a descrição e recuperação de informações na Web”, dissertação de mestrado defendido em 2004, o autor objetivou: investigar a informação na Web como objeto de estudo da CI e as tecnologias para descrição e recuperação de informações na Web que possibilitam a organização da informação digital, a partir dos seguintes objetivos específicos: analisar as linguagens e recursos tecno-metodológicos para descrição de informação na Internet; investigar a RI com foco nas propostas de novos padrões de estrutura de Dados; e identificar as novas ferramentas que vêm sendo discutidas e implementadas, que objetivam a organização da informação na Web. A WS aparece no trabalho como um cenário de desenvolvimento de tecnologias que podem propiciar melhorias quanto à OI digital, estreitando a relação da CI com atividades na Web.

O estudo da WS nas teses e dissertações na CI parece acontecer de maneira centrada na recuperação e representação da informação, numa perspectiva que a promove como um campo futuro e promissor para pesquisa na área. Vale salientar também que, a relação com a CC se estreita, não somente pelo uso de seus métodos e ferramentas, mas pela inserção dos seus profissionais e

pesquisadores.

Assim, parece necessário que outros aspectos sejam estudados: questões epistemológicas e sociais; a posição do usuário; etc., pois a WS, no que diz respeito às suas propostas e aos seus cenários, assim que ativada, pode ter grande impacto.

7 CONSIDERAÇÕES FINAIS

A CI, segundo Robredo (2003), tem como temas da pesquisa aplicada: arquitetura de sistemas aplicativos (sistemas de armazenagem e RI, dentre outros), arquitetura e gerenciamento de redes, automação de sistemas de informação; planejamento, implementação e operação de bibliotecas digitais; gestão automatizada de documentos; gestão automatizada de imagens e documentos multimídia; desenvolvimento de software, desenvolvimento e gestão de sistemas de inteligência; e outros. Essa proposta do autor, por si, já apresenta uma relação da CI com a WS, principalmente pela ênfase na RI, na representação da informação e do conhecimento, mesmo que não seja de forma direta (bibliotecas digitais), ou que seja apenas pelo uso dos padrões.

A Web, mesmo a sintática, já se baseava na idéia principal da CI, a RI por meio de associação de idéias, mas ela não possui estrutura para realizar uma representação adequada da informação, o que melhora a recuperação.

A WS resolve os problemas que impediam a Web se tornar um efetivo sistema de informação, mas ela caminhou pouco diante das propostas. Evidente que existe a necessidade de maturidade de usuários, dos desenvolvedores, dos pesquisadores e ferramentas para que isso aconteça. Entende-se que a CI deve, enquanto disciplina científica, colaborar com a WS. A pesquisa em pós-graduação no Brasil parece estar em rumo adequado no sentido de contribuir para o desenvolvimento da WS, focando principalmente a RI e a representação da informação e do conhecimento, campos em que área possui competências exclusivas.

Outras perspectivas que relacionam a CI e a WS foram identificadas. A primeira está inserida entre o contexto científico que embasa a CI, aprofundada pela explicação da interdisciplinaridade, que caracteriza a área e o papel da Web como ambiente principal de comunicação e de estoque de informação. Aceito ou não, o modelo científico chamado de pós-moderno existe e altera de maneira significativa a produção da informação e do conhecimento, ou seja, existe a necessidade de criar e adaptar instrumentos de representação. López-Huertas (2006) diz que o modelo disciplinar ainda hoje é a base para a criação de ferramentas de indexação, recuperação e organização da informação e conhecimento por parte da CI, criticando a não atualização de ferramentas da representação da informação, como a CDU e CDD.

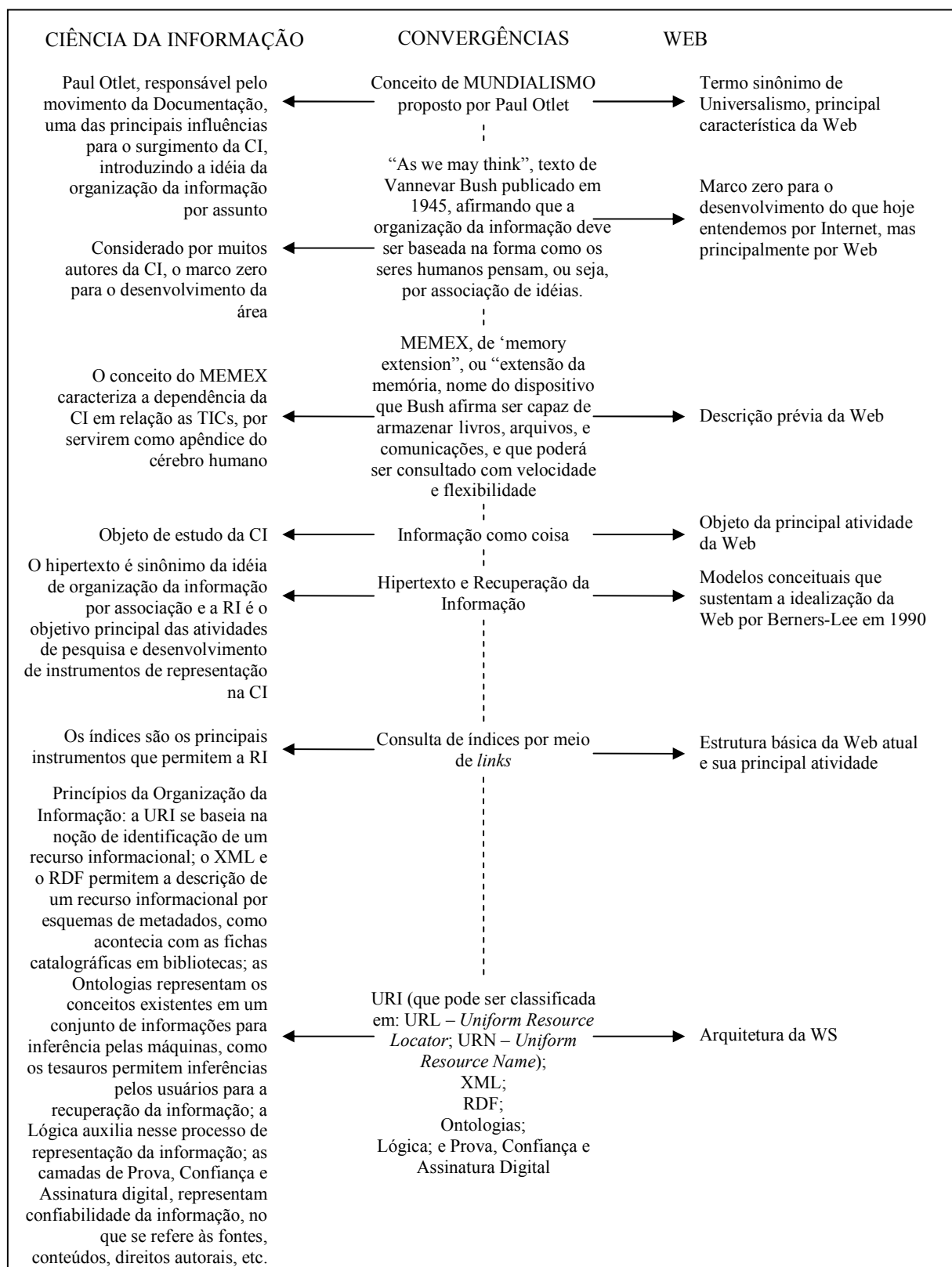
Esse exemplo se restringe à produção científico-tecnológica, que certamente é uma das áreas que produz os conjuntos mais estabilizados de registros, no que diz respeito aos formatos e padrões de

representação. A Web é um ambiente com todo tipo de informação, com todo tipo de influência, com todo tipo de usuário, devido à sua característica de universalidade. Se a informação científico-tecnológica necessita de propostas no que diz respeito à sua representação para recuperação, como ficam os outros tipos de informação que veiculam a Web, como sites pessoais, blogs, informação jornalística, comércio, etc.?

Entende-se, portanto, que esse aspecto deva ser um ponto de atuação entre a CI e esse ambiente de armazenamento do conhecimento objetivo: a Web, que na proposta da WS, procura evoluir para se tornar um efetivo sistema de informação e não apenas ser canal de comunicação entre os indivíduos. As mudanças na produção da ciência indicam a necessidade de uma representação da informação e do conhecimento mais personalizada e essa tendência deve ser levada a outras formas de gerar informação.

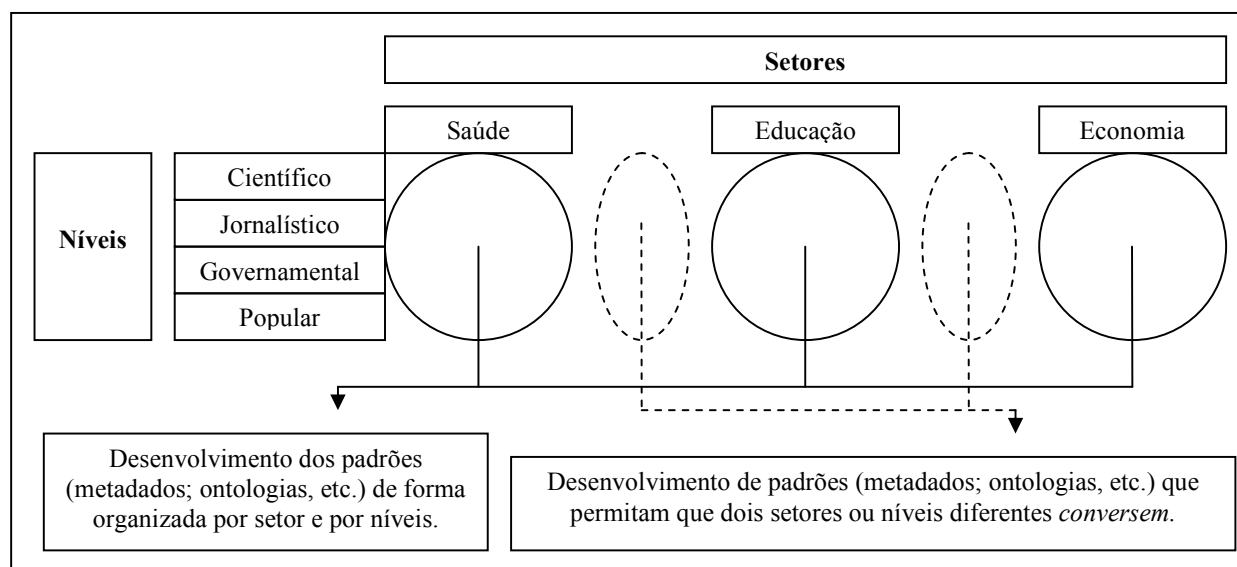
Um segundo aspecto que pode ser interesse de estudos entre a CI e a WS está relacionado com o quarto capítulo do trabalho: Tecnologia e Sociedade. Apesar da CI ter se originado, se desenvolver e depender das tecnologias, ela não deve deixar de avaliar os impactos sociais da tecnologia e as suas relações. Ou seja, a CI, assim como a sociedade em geral, deve utilizar as tecnologias como instrumentos, fazendo com que elas atuem visando atender as necessidades existentes e não de forma inversa. Nesse sentido, no que diz respeito à WS, a CI deve realizar pesquisas que levem em consideração principalmente as necessidades, os hábitos, as condições e outros fatores relacionados aos usuários. A WS, principalmente nos textos analisados nesse estudo, parece desconsiderar o usuário no seu planejamento e desenvolvimento. Assim, ela pode ‘falhar’ ou atrasar sua efetivação, se objetiva ser uma extensão da Web atual.

Outro aspecto que pode ser interesse de pesquisas entre a CI e a WS, diz respeito à origem da Web e da Internet e à noção da computação em rede. A CI, por muitos de seus autores, tem como seu marco zero o texto “As we may think” de Vannevar Bush e a Internet e Web também, assim como o conceito de ‘mundialismo’ proposto por Paul Otlet (MATTELART, 2006), que influenciam ambos os campos. Ou seja, uma ciência e uma tecnologia, que possuem a informação, como objeto principal, nascem da proposta de um mesmo autor, mas se desenvolvem separadas ou não tão ligadas como deveriam ter sido. A Figura 24 ilustra alguns pontos de convergências entre a CI e a Web identificados na presente pesquisa.

Figura 24. Algumas convergências entre a Ciência da Informação e a Web

A WS é uma realidade, talvez não no que diz respeito ao seu funcionamento efetivo, mas certamente como um vasto campo de pesquisa e de possibilidades de criação de tecnologias a ser desenvolvido e explorado. Mas apesar de seus objetivos ambiciosos, procura preservar as bases ideológicas da Web atual: o universalismo, a máxima descentralização possível e a tolerância. Evidentemente, que em certo sentido esses atributos podem ser prejudiciais para o seu desenvolvimento, pois essas características não dizem respeito somente ao uso do ambiente, mas principalmente ao seu desenvolvimento. Assim, a Web altera até mesmo o conceito de usuário, que aqui não é apenas aquele que utiliza algo, mas aquele que interfere na construção, que adota ou não padrões estabelecidos, que influencia comportamento, que forma opinião. Ou seja, a WS preserva as características mais inerentes da Web atual que se referem aos usuários, mas praticamente não fazem menção a eles no seu projeto. Nesse sentido, a Web 2.0 (social), apresentada no artigo “Hacia una web semántica social” de Morato et al. (2008), tem algumas lições a ensinar, como a valorização de um recurso da Web pelo seu uso (usabilidade alta) e pela arquitetura de colaboração. Essas características da Web social possibilitam entre outras coisas, que um recurso seja avaliado por diversos indivíduos, em diversos contextos. Algumas sugestões para o desenvolvimento da WS são: buscar um equilíbrio entre uma rigidez na utilização dos padrões, que permite a interoperabilidade, com a massificação do uso desses padrões, ou seja, envolver o usuário de Web na busca da popularização das tecnologias; trabalhar seu desenvolvimento por níveis e setores, conforme o exemplo na Figura 25 abaixo.

Figura 25. Esquema de sugestão para a criação de níveis e setores para o desenvolvimento da WS



A idéia representada pela Figura 25 apresenta apenas conjecturas, ou seja, os níveis e setores podem não ser os exclusivamente indicados e essa pode não ser a única opção para o desenvolvimento da WS, mas seguindo esse caminho a WS pode encontrar facilidade na sua efetivação. Exemplo: estabelecida as ontologias das áreas de saúde e de educação no nível científico, pode ficar mais fácil a criação de outra ontologia que permita o intercâmbio entre ambas, pois o uso isolado das ontologias não caracteriza uma interação semântica. Essa proposta também pode facilitar uma maior inserção dos usuários, no que se refere à usabilidade dos padrões da WS, devido ao conhecimento da área e à especificidade do nível.

A criação isolada de ontologias foi um tema constante presente na análise das teses e dissertações na área de CI com quatro dos trabalhos. Mas criação das ontologias de forma isolada não implica na efetivação da WS. Por isso, parece correto afirmar que muitos dos trabalhos utilizaram a WS apenas como um contexto onde as ontologias estavam inseridas, assim, não há uma contribuição direta para o seu desenvolvimento, no que diz respeito a um uso macro ?? (voce quis dizer amplo??) das ontologias.

Os atributos que mais caracterizam a Web (universalismo, a descentralização, a tolerância), além do ponto de vista do usuário, podem ser vistos de outra forma, já que como afirma Harmelen (2006) essa extensão da Web atual não é um campo de pesquisa somente da CC, tratando-se de um ambiente para todos os campos de pesquisa. A WS não é campo de pesquisa somente da CC por deliberação consciente dessa área, além do que essa ciência não dará conta de seu desenvolvimento de forma isolada. Assim o universalismo, a descentralização e a tolerância se estendem para a Web, na WS, no nível de pesquisa pura e aplicada, como objeto de estudo universal e descentralizado, no que diz respeito a todas as outras ciências.

Portanto, o desenvolvimento da WS é campo de pesquisa e deve ser preocupação primária da CI, considerando os pontos assinalados anteriormente: mundialismo, organização da informação por associação, TICs, informação como coisa, etc. Também são preocupações da CI visando uma participação efetiva na construção, aprimoramento e uso da WS: formação dos profissionais bibliotecários, arquivistas, museólogos; estudo de comunidades e usuários; dentre outros, enfim, temas para a pesquisa fundamental e para a aplicada da CI.

Os primeiros passos nesse sentido estão presentes principalmente nas teses e dissertações analisadas, que, no entanto, ainda não oferecem uma contribuição real para o desenvolvimento da WS, uma vez que ela é considerada o contexto para a discussão de metadados, ontologias, interoperabilidade, etc. Mas em relação à RI e à representação da informação e do conhecimento, há mais avanços, e a análise das teses e dissertações evidenciam isso.

Na análise das teses e dissertações, destaca-se o número alto de indivíduos oriundos da área da computação presentes nos cursos de pós-graduação da área de CI, o que reforça a relação interdisciplinar proposta. Recomenda-se que se dê maior atenção aos objetivos dos trabalhos de pós-graduação, no sentido de sua contribuição para o enriquecimento da área de CI, além de atender seus principais focos de pesquisa, como: representação da informação, no que se refere aos processos documentários, como a elaboração de linguagens documentárias; recuperação da informação, no que diz respeito à figura do usuário, criando ambientes onde sua visão e necessidade sejam as balizas.

As pesquisas realizadas apresentam como lacuna a falta de exposição das mesmas para as outras comunidades envolvidas no desenvolvimento da WS. Pinheiro (2008) demonstra que nas citações de teses e dissertações no Brasil sobre a WS, que relacionam os campos disciplinares da CI e da CC, a primeira cita 70% de trabalhos da CC, enquanto a última não cita nenhum da CI. Evidentemente, que são ciências com naturezas distintas, mas a distância deve diminuir, tendo em vista o potencial da área de CI para o alcance dos objetivos propostos para a WS. A partir, portanto, dessa constatação, indicam-se algumas sugestões:

- O estudo de citações relacionando a CI e a CC, em periódicos e eventos científicos, para verificar se o comportamento apresentado por Pinheiro (2008) com relação às teses e dissertações se repete;
- Indicação de como a WS se integra à literatura da CI no Brasil, através dos artigos de periódicos e dos trabalhos apresentados em anais de eventos, com o intuito de identificar a abrangência do estudo desse tema;
- Criação de uma secretaria para assuntos relacionados à pesquisa na WS na CI, por parte da Associação Nacional de Pesquisa e Pós-Graduação em CI (ANCIB), que represente a área junto ao W3C, que teve escritório inaugurado no Brasil em 2008, para que a produção de conhecimentos significativos sobre a Web, produzidos na área, seja aproveitada não somente por uma comunidade restrita e também para que alguns resultados de pesquisa aplicada possam ser utilizados, ou ao menos considerados como uma opção pela W3C; e

- Propostas curriculares nos cursos de Biblioteconomia e Ciência da Informação, que tenham disciplinas focada nessa relação CI e WS, que não é apenas tecnológica, mas que exige um trabalho intelectual de entendimento e de decisões quanto à representação documentária, relação com usuários, com domínios de conhecimento, etc. Ou seja, que prepare o profissional para contribuir no desenvolvimento efetivo da WS, como um conhecimento específico da área.

O trabalho de forma geral parece ter apresentado mais problemas e questões do que propriamente resultados, devido à abordagem exploratória adotada como método e da novidade do objeto estudado: a Web Semântica, cujos objetivos foram descritos a partir do ano de 2001. A pesquisa na CI sobre a WS, ainda por um tempo, parece que irá abordar o tema, de forma introdutória e ainda não tão delimitada, até que sejam definidas as fronteiras de atuação e colocadas em prática, de maneira ampla, as iniciativas concretas já criadas, como: a transição de tesauros para ontologias, conforme os textos de Pérez Agüera (2004), Guzmán Luna, Torres Pardo e López García (2006) e as ontologias criadas em alguns trabalhos de teses e dissertações da área. A CI tem se preocupado com WS e conforme demonstrado pelas teses e dissertações apresentadas o foco está na representação e recuperação da informação. Outros aspectos também merecem atenção, mas trata-se de um momento inicial. É evidente e clara a existência de espaço para a CI no desenvolvimento dessa tecnologia. No entanto, os resultados das pesquisas baseadas nos princípios norteadores da CI devem ultrapassar as fronteiras da área, para que sua contribuição resulte em aplicações tecnológicas para tornar a WS uma realidade.

REFERÊNCIAS

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v.32, n.3, p.7-20, set./dez. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19019.pdf>>. Acesso em: 14 set. 2007.

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da informação**, Florianópolis, n.15, 1º sem, 2003. Disponível em: <http://www.encontros-bibli.ufsc.br/Edicao_15/alvarenga_representacao.pdf>. Acesso: 10 mar. 2007.

ALVES, M. das D. R.; SOUZA, M. I. F. Estudo de correspondência de elementos metadados: Dublin Core e MARC 21. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 4, n. 2, p. 20-38, jan./jun. 2007. Disponível em: <<http://polaris.bc.unicamp.br/seer/ojs/include/getdoc.php?id=389&article=94&mode=pdf>>. Acesso em: 20 mar. 2008.

ALVITE DÍEZ, M. L. et al. **Docutes**: tesouro de Ciencias de la Documentación. León: Universidad de León, [2004]. Disponível em: <<http://www3.unileon.es/dp/abd/tesouro/pagina/tesdocumentacion/docutes.htm>>. Acesso em: 20 mar. 2008.

ARANO, S. Los tesauros y las ontologías en la Biblioteconomía y la Documentación. **Hipertext.Net**, n.3, 2005. Disponível em: <<http://www.hipertext.net/web/pag260.htm>>. Acesso em: 14 set. 2007.

ARAÚJO, C. A. Á. A Ciência da Informação como ciência social. **Ciência da Informação**, Brasília, v.32, n.3, p.21-27, set./dez. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19020.pdf>>. Acesso em: 20 mar. 2007.

ASSOCIAÇÃO para a promoção e desenvolvimento da Sociedade da Informação – ApdSI. **Glossário da Sociedade da Informação**: versão 2007. Caparica: ApdSI, 2007. Disponível em: <<http://www.anacom.pt/render.jsp?contentId=504232>>. Acesso em: 04 abr. 2008.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, [2000]. 225p.

BARRETO, A. de A. A condição da informação. **São Paulo em Perspectiva**, São Paulo, v.16, n.3, p.67-74, jul./set. 2002. Disponível em: <<http://www.scielo.br/pdf/spp/v16n3/13563.pdf>>. Acesso em: 20 mar. 2007.

BARRETO, A. de A. A eficiência técnica e econômica e a viabilidade de produtos e serviços de informação. **Ciência da Informação**, Brasília, v.25, n.3, p.405-414, set./dez.1996. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/466/425>>. Acesso em: 20 mar. 2007.

BARRETO, A. de A. A questão da informação. **São Paulo em Perspectiva**, São Paulo, v.8, n.4, p.3-8, out./dez. 1994. Disponível em: <<http://aldoibct.bighost.com.br/quest/quest2.pdf>>. Acesso em: 20 mar. 2007.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, May 2001. Disponível em: <http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 19 out. 2006.

BICALHO, L. M.; BORGES, M. E. N. Transdisciplinaridade na Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: Informação, conhecimento e

transdisciplinaridade – ENANCIB, 5., 2003, Belo Horizonte. **Anais...** Belo Horizonte: ENANCIB, 2003. 1 CD.

BORGES, M. A. G. A compreensão da sociedade da informação. **Ciência da Informação**, Brasília, v.29, n. 3, p.25-32, set./dez.2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n3/a03v29n3.pdf>>. Acesso em: 20 maio 2007.

BORKO, H. Information Science: what is it? **American Documentation**, Washington, p.3-5, Jan. 1968.

BRANDÃO, N. H. A interdisciplinaridade da biblioteconomia. **Boletim ABDF**, Nova Série, Brasília, v.5, n.4, p.21-44, out./dez. 1982.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: Diversidade cultural e políticas de informação – ENANCIB, 9., 2008, São Paulo. **Anais...** São Paulo: ENANCIB, 2008. 1 CD.

BREITMAN, K. K. **Web semântica: a Internet do futuro**. Rio de Janeiro: LTC, [c2006]. 190p.

BUCKLAND, M. K. **A brief biography of Suzanne Renée Briet**. 2005. Disponível em: <<http://people.ischool.berkeley.edu/~buckland/Brietaut2.pdf>>. Acesso em: 30 abr. 2007.

BUCKLAND, M. K. Descrição e pesquisa: metadados como infra-estrutura. **Brazilian Journal of Information Science – BJIS**, Marília, v.0, n.0, p.3-15, jul./dez. 2006. Disponível em: <<http://www.bjis.unesp.br/pt/viewarticle.php?id=3&layout=abstract>>. Acesso em: 14 mar. 2008.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science (JASIS)**, Washington, v.45, n.5, p.351-360, 1991.

BUSH, V. As we may think. **Atlantic Monthly**, Boston, v.176, n.1, p.101-108, 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em: 28 fev. 2004.

CABRÉ, M. T. La terminología hoy: concepciones, tendencias y aplicaciones. **Ciência da Informação**, Brasília, v.24, n.3, p.289-298, set./dez. 1995. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/487/442>>. Acesso em: 20 nov. 2008.

CABRÉ, M. T. La terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro. **Revista Debate Terminológico**, Paris, n.1, 2005. Disponível em: <http://www.riterm.net/revista/n_1/cabre.pdf>. Acesso em: 08 jun. 2008.

CABRERA GONZÁLEZ, M. A. La era de la participación y el creciente poder de las audiencias: (web 2.0 o web social y periodismo 3.0 o periodismo participativo). In: TOUTAIN, L. M. B. B. (Org.). **Para entender a Ciência da Informação**. Salvador: EDUFBA, 2007. p.163-184. (Saladeaula; v.6)

CAMPOS, M. L. M.; CAMPOS, M. L. de A.; CAMPOS, L. M. Web semântica e a gestão de conteúdos informacionais. In: KURAMOTO, H. (Org.). **Bibliotecas digitais: saberes e práticas**. 2.ed. Salvador: UFBA; Brasília: IBICT, 2006. p.55-77.

CAPURRO, R. Epistemologia e Ciência da Informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 5., 2003, Belo Horizonte. **Anais...** Belo Horizonte: ENANCIB, 2003. 1 CD.

CAREGNATO, R. C. A.; MUTTI, R. Pesquisa qualitativa: análise do discurso *versus* análise de conteúdo. **Texto & Contexto – Enfermagem**, Florianópolis, v.15, n.4, out./dez. 2006. Disponível em: <<http://redalyc.uaemex.mx/redalyc/pdf/714/71415417.pdf>>. Acesso em: 20 nov. 2008.

CHARTIER, R. Línguas e leituras no mundo digital. In: _____. **Os desafios da escrita**. São Paulo: UNESP, 2002a. p.11-32.

CHARTIER, R. Morte ou transfiguração do leitor? In: _____. **Os desafios da escrita**. São Paulo: UNESP, 2002b. p.101-124.

COSTA, C. I. da. Glossário. In: LÉVY, P. **Cibercultura**. 2.ed. São Paulo: Editora 34, 2003. (TRANS)

CUPANI, A. A tecnologia como problema filosófico: três enfoques. **Scientiae Studia**, São Paulo, v.2, n.4, p.493-518, 2004. Disponível em: <http://www.scientiaestudia.org.br/revista/PDF/02_04_02_Cupani.pdf>. Acesso em: 13 maio 2007.

DAGNINO, R.; BRANDÃO, F. C.; NOVAES, H. T. Sobre o marco analítico-conceitual da tecnologia social. In: **TECNOLOGIA social: uma estratégia para o desenvolvimento**. Rio de Janeiro: Fundação Banco do Brasil, 2004. p.15-64. Disponível em: <www.oei.es/salactsi/Teconologiasocial.pdf>. Acesso em: 12 abr. 2008.

DEMO, P. Elementos da metodologia dialética. In: _____. **Introdução à metodologia da ciência**. São Paulo: Atlas, 1983. p.85-100.

DENCKER, A. de F. M.; VIÁ, S. C. da. **Pesquisa empírica em ciências humanas**: (com ênfase em comunicação). 2.ed. São Paulo: Futura, 2002.

DENNING, P. J. et al. Computing as a discipline. **Communications of the ACM**, New York, v.32, n.1, p.9-23, Jan. 1989. Disponível em: <<http://cs.gmu.edu/cne/pjd/GP/CompDisc.pdf>>. Acesso em 12 maio 2008.

DENNING, P. J. Is Computer Science science? **Communications of the ACM**, New York, n.48, n.4, p.27-31, Apr. 2005. Disponível em: <<http://cs.gmu.edu/cne/pjd/PUBS/CACMcols/cacmApr05.pdf>>. Acesso em: 12 maio 2008.

DUBOIS, J. et al. **Dicionário de lingüística**. São Paulo: Cultrix, [1999]. 653p.

DZIEKANIAK, G. V.; KIRINUS, J. B. Web semântica. **Encontros Biblio: R. Eletr. Bibliotecon. Ci. Inf.**, Florianópolis, n.18, 2º sem. 2004. Disponível em: <http://www.encontros-bibli.ufsc.br/Edicao_18/2_Web_Semantica.pdf>. Acesso em: 25 mar. 2007.

FERREIRA, A. B. de H. **Novo dicionário eletrônico Aurélio versão 5.11**. 2004. 1 CD.

FOSKETT, A. C. **A abordagem temática da informação**. São Paulo: Polígono, 1973. 437p.

FRANCELIN, M. M. A epistemologia da complexidade e a Ciência da Informação. **Ciência da Informação**, Brasília, v.32, n.2, p.64-68, maio/ago. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n2/17034.pdf>>. Acesso em: 17 abr. 2004.

FRANCELIN, M. M. Configuração epistemológica da Ciência da Informação no Brasil em uma perspectiva pós-moderna: análise de periódicos da área. **Ciência da Informação**, Brasília, v.33, n.2, p.49-66, maio/ago. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a05v33n2.pdf>>. Acesso em: 24 set. 2005.

FREIRE, G. H. Ciência da Informação: temática, histórias e fundamentos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.11, n.1, p.6-19, jan./abr. 2006. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/442/253>>. Acesso em: 12 abr. 2007.

FREIRE, I. M. A utopia planetária de Pierre Lévy: uma leitura hipertextual d'a inteligência coletiva. **Perspectiva em Ciência da Informação**, Belo Horizonte, v.10, n.2, p.132-139, jul./dez. 2005. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/download/340/147>>. Acesso em: 12 abr. 2007.

GADOTTI, M. A dialética: concepção e método. In: _____. **Concepção dialética da educação**: um estudo introdutório. 3.ed. São Paulo: Autores Associados; Cortez, 1984. p.15-38. (Educação Contemporânea)

GIL, A. C. **Como elaborar projetos de pesquisa**. 4.ed. São Paulo: Atlas, 2002. 159p.

GONZÁLEZ DE GÓMEZ, M. N. et al. Quem é o sujeito da pesquisa inter e trans-disciplinar: buscando desenvolver um modelo de análise. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: Informação, conhecimento e transdisciplinaridade – ENANCIB, 5., 2003, Belo Horizonte. **Anais...** Belo Horizonte: ENANCIB, 2003.

GUIMARÃES E SILVA, J. Ciência da Informação: uma ciência do paradigma emergente. In: PINHEIRO, L. V. R. (Org.). **Ciência da Informação, Ciências Sociais e interdisciplinaridade**. Brasília: IBICT, 1999. p.79-117.

GUINCHAT, C.; MENO, M. **Introducción general a las ciencias y técnicas de la información y de la documentación**. París: UNESCO, 1983. 480p. Disponível em: <<http://unesdoc.unesco.org/images/0004/000473/047393so.pdf>>. Acesso em: 14 set. 2007.

GUSMÁN LUNA, J.; TORRES PARDO, D.; LÓPEZ GARCÍA, A. N. Desarrollo de una ontología en el contexto de la web semántica a partir de un tesoro documental tradicional. **Rev. Interam. Bibliot.**, Medellín, v.29, n.2, jul./dic. 2006.

HARMELEN, F. v. **Where does it break?** or: why Semantic Web research is not just “Computer Science as usual”. [2006]. Disponível em: <www.eswc2006.org/keynote-frank-van-harmelen.pdf>. Acesso em: 20 mar. 2008.

HJORLAND, B. **Core concepts in Library and Information Science (LIS)**. c2005. Disponível em: <<http://www.db.dk/bh/core%20concepts%20in%20lis/home.htm>>. Acesso em: 10 jul. 2008.

HJORLAND, B. **Lifeboat for Knowledge Organization**. [2008]. Disponível em: <http://www.db.dk/bh/lifeboat_ko/List%20of%20concepts.htm>. Acesso em 10 nov. 2008.

INGWERSEN, P. Conceptions of Information Science. In: VAKKARI, P. RONIN, B. (ed.) **Conceptions of Library and Information Science**: historical, empirical and theoretical perspectives. London: Taylor Graham, 1992. p.299-312.

JACOB, E. K. Ontologies and the semantic web. **Bulletin of the American Society for Information Science and Technology**, Silver Spring, p. 16-18, Apr./May. 2003. Disponível em: <<http://www.asis.org/Bulletin/Apr-03/jacob.html>>. Acesso em: 10 mar. 2007.

JAPIASSU, H. **Interdisciplinaridade e patologia do saber**. Rio de Janeiro: Imago, 1976. 220p. (Logoteca)

KERCKHOVE, D. de. Texto, contexto e hipertexto: três condições da linguagem, três condições da mente. **Revista FAMECOS**, Porto Alegre, n.22, dez. 2003. Disponível em: <<http://www.pucrs.br/famecos/pos/revfamecos/22/a01v1n22.pdf>>. Acesso em: 15 fev. 2008.

- KONDER, L. **O que é dialética?** 28.ed. São Paulo: Brasiliense, 1998. 87p. (Primeiros Passos; v.23)
- KUHN, T. S. **A estrutura das revoluções científicas.** 4.ed. São Paulo: Perspectiva, 1996. 257p. (Debates: Ciências; v.115)
- LASSILA, O.; SWICK, R. R. **Resource Description Framework (RDF): model and syntax specification.** 1999. Disponível em: <www.w3.org/TR/1999/REC-rdf-syntax-19990222/>. Acesso em: 20 mar. 2008.
- LEÃO, L. **O labirinto da hipermídia:** arquitetura e navegação no ciberespaço. 2.ed. São Paulo: FAPESP, Iluminuras, 2001. 158p.
- LE COADIC, Y.-F. **A Ciência da Informação.** Brasília: Briquet de Lemos Livros, 1996. 119p.
- LE COADIC, Y.-F. A matemática da informação. In: TOUTAIN, L. M. B. B. (Org.). **Para entender a Ciência da Informação.** Salvador: EDUFBA, 2007. p.219-239. (Sala de aula; v.6)
- LÉVY, P. **A inteligência coletiva:** por uma antropologia do ciberespaço. 2.ed. São Paulo: Edições Loyola, 1999. 212p.
- LÉVY, P. **Cibercultura.** 2.ed. São Paulo: Editora 34, 2003. 260p. (TRANS)
- LÉVY, P. O paradigma informático. In: _____. **A máquina universo:** criação, cognição e cultura informática. Porto Alegre: Artmed, 1998. p.103-118.
- LIMA, G. A. B. Interfaces entre a Ciência da Informação e a Ciência Cognitiva. **Ciência da Informação,** Brasília, v.32, n.1, p.77-87, jan./abr. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n1/15975.pdf>>. Acesso em: 06 maio 2008.
- LÓPEZ-HUERTAS, M. J. Análisis del dominio interdisciplinar para la representación y organización del conocimiento. In: GONZÁLEZ DE GÓMEZ, M. N.; ORRICO, E. G. D. **Políticas de memória e informação:** reflexos na organização do conhecimento. Natal: EDUFRN, 2006. p.209-236.
- LUÍS GARCIA, J. Sobre as origens da crítica da tecnologia na teoria social: Georg Simmel e a autonomia da tecnologia. **Scientiae Studia,** São Paulo, v.5, n.3, p.287-336, 2007. Disponível em: <http://www.scientiaestudia.org.br/revista/PDF/05_03_02.pdf>. Acesso em: 03 maio 2008.
- MANIEZ, J. **Los lenguajes documentales y de clasificación:** concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación Germán Sánchez Ruipérez, 1993. 230p. (Biblioteca del Libro; v.56)
- MARCONI, M. de A.; LAKATOS, E. M. **Metodologia do trabalho científico.** 6.ed. rev. e ampl. São Paulo: Atlas, 2001. 219p.
- MATTELART, A. **História da sociedade da informação.** 2.ed. rev. e atual. São Paulo: Edições Loyola, 2006. 197p.
- MEADOW, C. T.; BOYCE, B. R.; KRAFT, D. H. **Text information retrieval systems.** 2.ed. San Diego: Academic Press, 2000. 364p.
- MEIRELLES, F. de S. **Informática:** novas aplicações com microcomputadores. 2. ed., atual. e ampl. São Paulo: Makron Books, 1994. 615p.

MÉNDEZ RODRÍGUEZ, E. **Metadatos y recuperación de información**: estándares, problemas y aplicabilidad en bibliotecas digitales. Gijón: Trea, 2002. 429p. (Biblioteconomía y Administración Cultural; 66)

MENDONÇA, A. V. M. O uso da análise do discurso do sujeito coletivo em Ciência da Informação. In: MUELLER, S. P. M. (Org.). **Métodos para a pesquisa em Ciência da Informação**. Brasília: Thesaurus, 2007. p.149-170. (Ciência da Informação e Comunicação)

MENDONÇA, E. S. A lingüística e a ciência da informação: estudos de uma interseção. **Ciência da Informação**, Brasília, v.29, n.3, p.50-70, set./dez. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n3/a06v29n3.pdf>>. Acesso em: 20 maio 2008.

MILLER, E. **Digital Libraries and the Semantic Web**. 2001. Disponível em: <<http://www.w3.org/2001/09/06-ecdl/slide17-0.html>>. Acesso em: 20 mar. 2008.

MIRANDA, A. A Ciência da Informação e a teoria do conhecimento objetivo: um relacionamento necessário. In: _____. **Ciência da Informação**: teoria e metodologia de uma área em expansão. Brasília: Thesaurus, 2003. p.173-188.

MIRANDA, A. Sociedade da informação: globalização, identidade cultural e conteúdos. **Ciência da Informação**, Brasília, v.29, n.2, p.78-88, maio/ago. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n2/a10v29n2.pdf>>. Acesso em: 20 set. 2006.

MIRANDA, A. S. S. **Ontologias**: indexação e recuperação de fotografias baseadas na técnica fotográfica e no conteúdo da imagem. 2007. 130f. Dissertação (Mestrado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília. 2007. Disponível em: <<http://repositorio.bce.unb.br/handle/123456789/1013>>. Acesso em: 20 mar. 2008.

MORAIS, E. F.; SOARES, M. B. **Web semântica para máquinas de busca**. 2002. Disponível em: <<http://homepages.dcc.ufmg.br/~nivio/cursos/pa03/seminarios/seminario7/seminario7.pdf>>. Acesso em: 04 out. 2006.

MORAIS, R. de. A realidade tecnológica. In: _____. **Filosofia da ciência e da tecnologia**: introdução metodológica e crítica. 5.ed. Campinas: Papirus, 1988. p.99-156.

MORATO, J. et al. Hacia una web semántica social. **El Profesional de la Información**, Barcelona, v.17, n.1, p.78-85, Jan./Feb. 2008.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A. de P. Thesaurus and Ontology: a study of the definitions found in the Computer and Information Science Literature, by means of an analytical-synthetic method. **Knowledge Organization**, Wurzburg, v.31, n.4, p.231-244, 2004.

MORIN, E. Epistemologia da tecnologia. In: _____. **Ciência com consciência**. 7.ed. Rio de Janeiro: Bertrand Brasil, 2003a. p.107-115.

MORIN, E. Por uma reforma do pensamento. In: PENA-VEGA, A.; NASCIMENTO, E. P. do (Orgs.). **O pensar complexo**: Edgar Morin e a crise da modernidade. 2.ed. Rio de Janeiro: Garamond, 1999. p.21-34.

MORIN, E. A responsabilidade do pesquisador perante a sociedade e o homem. In: _____. **Ciência com consciência**. 7.ed. Rio de Janeiro: Bertrand Brasil, 2003b. p.117-123.

MOSTAFA, S. P. **Epistemologia da Biblioteconomia**. São Paulo, 1985. 145f. Tese (Doutorado) – Curso de Pós-Graduação em Filosofia da Educação, Pontifícia Universidade Católica de São Paulo. 1985.

NAZARENO, C. et al. **Tecnologias da informação e sociedade: o panorama brasileiro**. Brasília: Plenarium, 2007. 187p. Disponível em: <http://www.camara.gov.br/internet/infdoc/Publicacoes/html/pdf/tecnologia_info.pdf>. Acesso em: 12 maio 2008.

OLIVEIRA, E. C. **Autoria de documentos para a web semântica: um ambiente de produção de conhecimento baseado em ontologias**. 2006. 207f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação e Documentação, Universidade de Brasília. 2006. Disponível em: <http://bdtd.bce.unb.br/tesesimplificado/tde_busca/arquivo.php?codArquivo=1565>. Acesso em: 20 mar. 2008.

ORRICO, E. G. D. Interdisciplinaridade: Ciência da Informação & Lingüística. In: PINHEIRO, L. V. R. (Org.). **Ciência da Informação, Ciências Sociais e interdisciplinaridade**. Brasília: IBICT, 1999. p.143-154.

ORTEGA, C. D. **Informática documentária: estado da arte**. São Paulo, 2002, 259f. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo. 2002. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27143/tde-27032004-155935/publico/Ortega.pdf>>. Acesso em: 12 maio 2008.

OTLET, P. O livro e a medida: Bibliometria. In: FONSECA, E. N. (Org.). **Bibliometria: teoria e prática**. São Paulo: Cultrix; USP, 1986. p.19-34.

PÉREZ AGÜERA, J. R. Automatización de tesauros y su utilización en la web semântica. **BID: Textos Universitaris de Biblioteconomia i Documentació**, Barcelona, v.13, des. 2004. Disponível em: <<http://www.ub.es/bid/13perez2.htm>>. Acesso em: 14 nov. 2008.

PINHEIRO, C. B. F. **A construção do conhecimento científico: a Web Semântica como objeto de estudo**. 2008. 63f. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2008. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=101422>. Acesso em: 08 dez. 2008.

PINHEIRO, L. V. R. Campo interdisciplinar da Ciência da Informação: fronteiras remotas e recentes. In: _____. (Org.). **Ciência da Informação, Ciências Sociais e interdisciplinaridade**. Brasília: IBICT, 1999. p.155-182.

POPPER, K. R. **Conhecimento objetivo: uma abordagem evolucionária**. Belo Horizonte: Itatiaia; São Paulo: USP, 1975. 394p. (Espírito no Nosso Tempo; v.13)

RAYWARD, W.B. The origins of Information Science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). **Journal of the American Society for Information Science (JASIS)**, Washington, v.48, n.4, p.289-300, 1997.

REITZ, J. M. **ODLIS: Online Dictionary for Library and Information Science**. 2007. Disponível em: <<http://lu.com/odlis/>>. Acesso em: 19 jun. 2008.

ROBREDO, J. **Da Ciência da Informação revisitada aos sistemas humanos de informação**. Brasília: Thesaurus; SSRRI Informações, 2003. 245p.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas**. 4.ed. ver. e ampl. Brasília: [Autor], 2005.

ROCHA, D.; DEUSDARÁ, B. Análise de conteúdo e análise do discurso: aproximações e afastamentos da (re)construção de uma trajetória. **ALEA: Estudos Neolatinos**, Rio de Janeiro, v.7, n.2, p.305-322, jul./dez. 2005. Disponível em: <<http://www.scielo.br/pdf/alea/v7n2/a10v7n2.pdf>>. Acesso em 26 abr. 2008.

SANTAELLA, L. **Matrizes da linguagem e pensamento**: sonora, visual, verbal: aplicações na hipermidia. São Paulo: FAPESP; Iluminuras, 2001. 432p.

SANTOS, B. de S. **Um discurso sobre as ciências**. Lisboa: Afrontamento, 1993. 58p.

SANTOS, D. P. L. dos. **Um arcabouço teórico para autoria de documentos visando atenuar o surgimento do fenômeno da ambigüidade**. Brasília, 2006, 181f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Departamento de Ciência da Informação e Documentação, Universidade de Brasília. 2006. Disponível em: <http://bdtd.bce.unb.br/tesesimplificado/tde_busca/arquivo.php?codArquivo=1029>. Acesso em: 12 maio 2008.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.1, p.41-62, jan./jun. 1996. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/235/22>>. Acesso em: 26 maio 2007.

SEVERINO, A. J. Metodologia do trabalho científico. 22.ed. rev. e ampl. São Paulo: Cortez, 2003. 335p.

SHERA, J. H. Sobre biblioteconomia, documentação e ciência da informação. In: GOMES, H. E. (Org.). **Ciência da Informação ou Informática?** Rio de Janeiro: Calunga, 1980. p.91-105. (Ciência da Informação)

SILVA, R. J. da. Faces da pesquisa e da interdisciplinaridade em Ciência da Informação no Brasil. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO: A política científica e os desafios da sociedade do conhecimento – ENANCIB, 6., 2005, Florianópolis. **Anais...** Florianópolis: ENANCIB, 2005.

SILVA, R. R. G. de. Informação, ciberespaço e consciência. **TransInformação**, Campinas, v.18, n.3, p.191-202, set./dez. 2006. Disponível em: <<http://revista.ibict.br/pbcib/index.php/pbcib/article/view/552>>. Acesso em: 10 abr. 2007.

SLYPE, G. V. **Los lenguajes de indización**: concepción, construcción y utilización en los sistemas documentales. Madrid: Fundación Germán Sánchez Ruipérez, 1991. 198p. (Biblioteca del Libro)

SMIT, J. **O que é documentação**. São Paulo: Brasiliense, 1986. 83p. (Primeiros passos, v.174)

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, Brasília, v.29, n.1, p.93-102, jan./abr. 2000. Disponível em: <<http://www.scielo.br/pdf/ci/v29n1/v29n1a10.pdf>>. Acesso em: 14 set. 2007.

SOUZA, R. R.; ALVARENGA, L. A Web semântica e suas contribuições para a Ciência da Informação. **Ciência da Informação**, Brasília, v.33, n.1, p.132-141, jan./abr. 2004. Disponível em: <<http://www.scielo.br/pdf/ci/v33n1/v33n1a16.pdf>>. Acesso em: 25 nov. 2006.

STOKES, D. E. **O quadrante de Pasteur**: a ciência básica e a inovação tecnológica. Campinas: UNICAMP, 2005. 246p. (Clássicos da Inovação)

STUMPF, I. R.; WEBER, M. H. Comunicação e informação: conflitos e convergências. In: LOPES, M. I. **Epistemologia da comunicação**. São Paulo: Loyola, 2003. p.121-134.

TÁLAMO, M. de F. G. M.; LARA, M. L. G. de. O campo da Lingüística Documentária. **TransInformação**, Campinas, v.18, n.3, p.203-211, set./dez. 2006. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewarticle.php?id=182>>. Acesso em: 10 out. 2008.

TARGINO, M. das G. Novas tecnologias de comunicação: mitos, ritos ou ditos? **Ciência da Informação**, Brasília, v.24, n.2, p.194-203, maio/ago. 1995. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewFile/553/502>>. Acesso em: 16 maio 2007.

TEIXEIRA, J. de F. **O que é inteligência artificial**. São Paulo: Brasiliense, 1990. 76p. (Primeiros passos; 230)

TOLOSA, G. H.; BORDIGNON, F. R. A. **Introducción a la recuperación de información**: conceptos, modelos y algoritmos básicos. Buenos Aires: Universidad Nacional de Luján, [2007]. 149p. Disponível em: <<http://eprints.rclis.org/archive/00014562/01/Introduccion-RI-v9f.pdf>>. Acesso em: 14 set. 2008.

UNIVERSIDADE de Aveiro. **Os serviços da Internet**. [2003]. Disponível em: <http://sweet.ua.pt/~a35438/Enc_Inf_Web_sites/Enc_Inf_Web1.htm>. Acesso em: 10 out. 2008.

URBIZAGÁSTEGUI ALVARADO, R. A bibliometria: história, legitimação e estrutura. In: TOUTAIN, L. M. B. B. (Org.). **Para entender a Ciência da Informação**. Salvador: EDUFBA, 2007. p.185-217. (Saladeaula; v.6)

VALENTIM, M. L. P. Análise de conteúdo. In: ____ (Org.). **Métodos qualitativos de pesquisa em Ciência da Informação**. São Paulo: Polis, 2005. p.119-134. (Palavra-Chave; v.16)

WERSIG, G. Information Science: the study of postmodern knowledge usage. **Information Processing & Management**, Elmsford, v. 29, n. 2, p.229-239, 1993.

WORLD Wide Web Consortium. **Long description of W3C10 timeline graphic**. 2005. Disponível em: <<http://www.w3.org/2005/01/timelines/timeline-2500x998.png>>. Acesso em: 12 ago. 2007.

WORLD Wide Web Consortium. **Sobre o consórcio W3C**. 2008. Disponível em: <<http://www.w3c.br/sobre/#history-head>>. Acesso em: 08 nov. 2008.

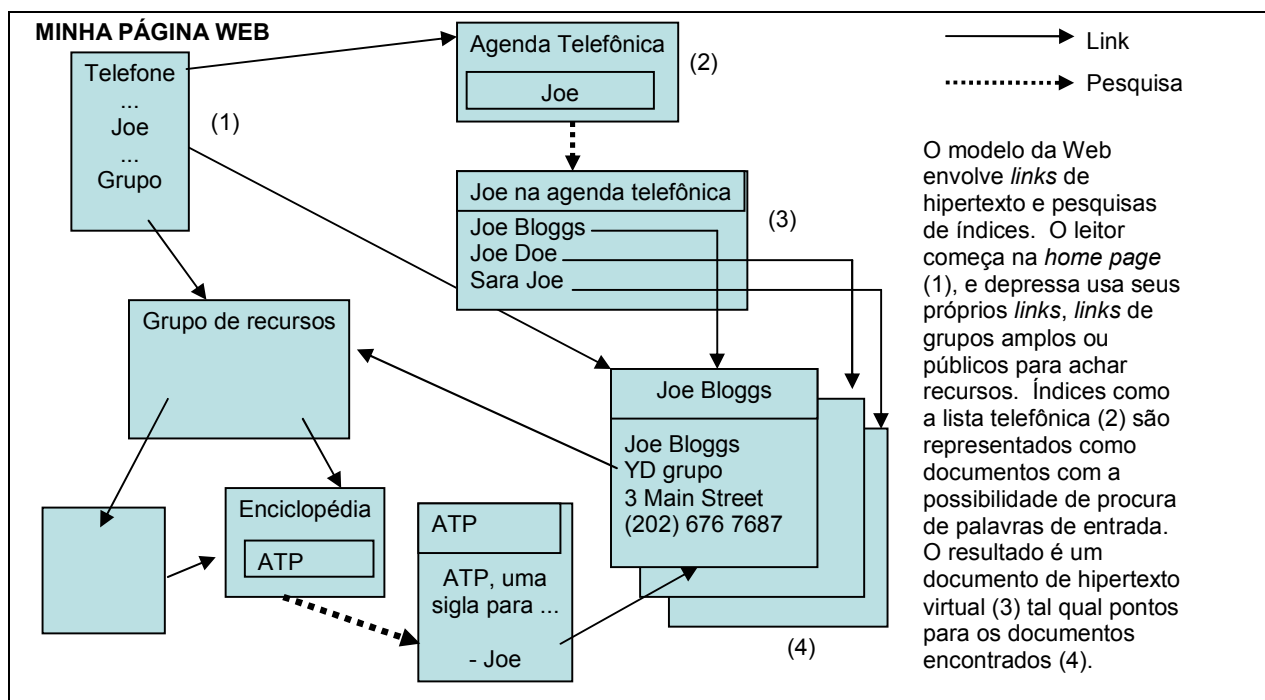
ZHONGHONG, W.; CHAUDHRY, A. S.; KHOO, C. Potential and prospects of taxonomies for content organization. **Knowledge Organization**, Wurzburg, v.33, n.3, p.160-169. 2006.

APÊNDICES

Apêndice A. Quadros da análise de conteúdo dos textos selecionados para identificação das propostas e desenvolvimento da Web Semântica.

QUADRO 1

Referência: BERNERS-LEE, T. et al. World-Wide Web : the information universe. 1992. Disponível em: <www.w3.org/History/1992/ENRAP/Article_9202.ps>. Acesso em: 10 dez. 2006.
<p>Conceito(s)/característica(s):</p> <ul style="list-style-type: none"> - A junção da técnica de hipertexto, recuperação da informação e o estabelecimento de redes de longo alcance. - Informações só precisam ser representadas uma vez. - A web se estende sem emendas como notas pessoais na área de trabalho local para bancos de dados grandes em outros locais. - Os documentos na Web não têm que existir como arquivos: eles podem ser documentos virtuais gerados por um servidor. - A estrutura em hipertexto permite mais escolhas que os sistemas de menus
<p>Objetivo(s):</p> <ul style="list-style-type: none"> - Recuperação de textos com dedução de idéias, através do conceito de hipertexto.
<p>Estrutura:</p> <ul style="list-style-type: none"> - Links: possibilitam uma topologia da informação que evoluem, enquanto modelam o estado do conhecimento humano. - Índices: são documentos, e assim podem ser encontradas através de buscas. Um índice é representado ao usuário por uma “página de cobertura” que descreve os dados indexados e as propriedades da máquina de procura. - Esquema comum nomeando documentos: sintaxe no estilo de SGML - Protocolos comuns de acesso de rede: HTTP - Formato comum de dados para hipertexto: WAIS
<p>Modelo(s) conceitual(is) utilizado(s):</p> <ul style="list-style-type: none"> - Hipertexto. - Recuperação da informação.
Ferramenta(s) tecnológica(s):
Cenário:



QUADRO 2

Referência: BERNERS-LEE, T. Web architecture from 50,000 feet . 1998. Disponível em: < http://www.w3.org/DesignIssues/Architecture.html >. Acesso em: 21 ago. 2008
Conceito(s)/característica(s): <ul style="list-style-type: none"> - A web é definida como um universo de informação em rede. - Simplicidade e mobilidade por parte da engenharia de software. - Descentralização e tolerância, princípios que regem a Internet. - A Web é por desígnio e filosofia um sistema descentralizado. - Um esquema novo de URI impede a interoperabilidade da Web. - Uma das diretrizes para desenvolvimentos da Web é que devem respeitar a definição genérica e sintaxe das URIs, não introduzindo um novo esquema de URI sem causa devida. - A melhora da URI envolve maturidade de ferramenta, educação de usuário e maturidade dos envolvidos com a Web. - Dificuldades na transição do HTML (que não tem espaço especial na arquitetura da Web, mas que ganhou domínio) para o HTML baseado no XML: <ul style="list-style-type: none"> - Não adesão de páginas na Web ao padrão HTML 4.0 ou SGML. - <i>Browsers</i> por muito tempo farão a leitura das páginas Web desenvolvidas dessa forma. - Muitos <i>browsers</i> não podem fazer a leitura do XML. - Necessidade de modelo generalista comum para a realização de inferências semânticas (WS) - Máquinas não podem responder perguntas arbitrárias.
Objetivo(s): <ul style="list-style-type: none"> - A interoperabilidade. - Universalidade de acesso, independente de hardware ou plataforma de software, infra-estrutura de rede, idioma, cultura, local geográfico, ou necessidade física ou mental. - Tornar a informação compreensível por máquinas: Web Semântica
Estrutura (Web Atual): <ul style="list-style-type: none"> - Universal Resource Identifier (URI): identificador universal de um recurso Web. - HTTP: um tipo de URI. Combina um protocolo básico de transporte com uma variedade limitada de metadados. - Domain Name System (DNS): Sistema de nome domínio (parte do HTTP), única parte da Web que é centralizada. - eXtensible Markup Language (XML):

- Resource Description Framework (RDF): Estrutura de descrição de recursos.
Estrutura (Web Semântica):
- XML: permite um formato básico para a estruturação de documentos.
- RDF: permite modelo básico de afirmação e cotação (resposta à necessidade de um modelo generalista de inferências semânticas).
- Camada ontológica: permite esquemas de conceitos mais poderosos.
- Camada de regras: uma linguagem de conversão possibilita a expressão de regras de conclusão, que permitem informação em um esquema ser deduzido de um documento em outro.
- Camada lógica: transformada em uma linguagem universal pode unificar todos os sistemas de dados, como o HTML era uma linguagem que unificava todos os sistemas de documentação humanos.
- Camada de prova: uma forma de RDF que permite para um agente enviar a outro uma afirmação, junto com o caminho de conclusão para aquela afirmação de suposições.
Modelo(s) conceitual(is) utilizado(s):
- Metadados
- Índice de condições
Ferramenta(s) tecnológica(s):
Cenário(s):

QUADRO 3

Referência: BERNERS-LEE, T. Semantic web road map . 1998. Disponível em: < www.w3.org/DesignIssues/Semantic.html >. Acesso em: 20 mar. 2007.
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - Um dos principais obstáculos para a efetivação da Web como sistema de informação é o fato de que a maior parte das informações na Web é projetada para consumo humano. - Necessidade de um modelo comum de grande generalidade para asserções semânticas: Resource Description Framework (RDF): Estrutura de descrição de dados. <p>Duas funções fundamentais os mecanismos de RDF devem ser capazes de fazer:</p> <ol style="list-style-type: none"> 1. que uma implementação versão 'n' seja capaz de ler esboços de RDF suficientes de modo a ser capaz de deduzir como ler um documento versão 'n+1'; 2. que uma aplicação do tipo A desenvolvida independentemente da aplicação de um tipo B, que tenha a mesma ou função similar, seja capaz de ler e processar esboços de informações suficientes para ser capaz de processar os dados da aplicação do tipo B. <ul style="list-style-type: none"> - Assinatura digital: os documentos serão analisados não só em árvores de asserções, mas em árvores de asserções sobre quem assinou quais asserções. A validação de prova, por regras de inferência, verificará a lógica, mas pelas asserções pelas quais um documento tenha sido assinado, verificará a assinatura. - A web caminha a partir dessas indicações para se tornar um grande base de dados.
<p>Objetivos:</p> <ul style="list-style-type: none"> - Roteiro para a modelagem da futura Web. - Realização de um conjunto de aplicações relacionadas a dados na Web. - A direção da Web Semântica desenvolve linguagens para expressar a informação num formulário processável por máquinas. - As aplicações que rodam por toda a web devem ser capazes de usar uma estrutura comum para combinar as informações de todas estas aplicações
<p>Estrutura:</p> <ul style="list-style-type: none"> - Linguagem de conversão: um exemplo simples da aplicação desta camada é quando duas bases de dados, construídas de forma independente e colocadas na web, estão ligadas por relações semânticas que permitem que as consultas em uma sejam convertidas em consultas da outra. - Camada Lógica: é preciso resolver o (limitado) poder do mecanismo de raciocínio que o receptor deve ter, e definir um subconjunto pleno de RDF que deverá ser compreendido; provavelmente se definirão algumas funções abreviadas para transmitir eficazmente as expressões dentro do conjunto de documentos no âmbito da linguagem restrita. - Validação de Provas: Os documentos enviados pela Internet serão escritos numa linguagem completa. Contudo, terão de ser restringidos de forma que, no caso de buscas, os resultados sejam computáveis, e na maioria dos casos, sejam provas. - Linguagens de consultas: RDF e buscadores de índices.
<p>Modelos conceituais utilizados:</p> <ul style="list-style-type: none"> - RDF (camada de esboço)

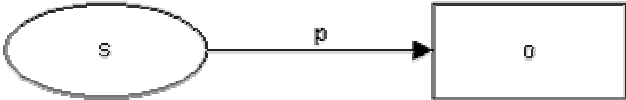
Ferramentas tecnológicas:
Cenários:

QUADRO 4

Referência: BERNERS-LEE, T. What the semantic web can represent . 1998. Disponível: < www.w3.org/DesignIssues/RDFnot.html >. Acesso em: 20 mar. 2007.
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - Questionamentos sobre a impossibilidade das metas da WS. - O conceito de documentos compreensíveis por máquinas, não insinua que a WS desenvolva ‘mágica’ para que as máquinas compreendam os murmúrios dos seres humanos. - O conceito de documentos compreensíveis por máquinas, indica apenas que as máquinas resolverão problemas bem definidos, através de operações bem definidas de dados. - O RDF se apresenta como uma linguagem completa para resolver problemas nesses níveis. - Aplicações que geram RDF serão limitados a gerar expressões simples, como listas de controle de acesso, preferências de privacidade, e critérios de busca. - Enquanto bancos de dados relacionais tratam com informações em uma especialidade, a WS visa unir dados de muitos modelos diferentes. - A WS precisa apresentar garantias. - O modelos precisa fazer relações automática, por exemplo: ele precisa ‘saber’ que a filha da filha é a neta, ou que o pai do filho é o avô.
<p>Objetivos:</p> <ul style="list-style-type: none"> - Via global para a representação do conhecimento.
Estrutura:
<p>Modelos conceituais utilizados:</p> <ul style="list-style-type: none"> - RDF: em certo sentido é um modelo de entidades e relações, mas as relações são em um primeiro momento objetos de classes, identificados por uma URI. Também não é modelo de conclusão.
Ferramentas tecnológicas:
Cenários:

QUADRO 5

Referência: LASSILA, O.; SWICK, R. R. Resource Description Framework (RDF): model and syntax specification . 1999. Disponível em: < www.w3.org/TR/1999/REC-rdf-syntax-19990222/ >. Acesso em: 20 mar. 2008.
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - RDF é um modelo de metadados. - Necessita de um modelo de especificação de classes: esquema. - O RDF se materializa usando a sintaxe XML - Além de declarações sobre recursos da Web, o RDF pode fazer declarações de outras declarações RDF. - Propriedades para modelos de declarações em RDF: <ul style="list-style-type: none"> - Assunto: identifica o recurso que é descrito pela declaração. - Predicado: identifica a propriedade original na declaração. - Objeto: identifica o valor de propriedade na declaração. - Tipo: descreve o tipo do recurso.
<p>Objetivos:</p> <ul style="list-style-type: none"> - RDF objetiva fazer uso (processar) de metadados. - RDF objetiva tornar possível especificar semanticamente um banco de dados em XML, permitindo interoperabilidade.
<p>Estrutura:</p> <ul style="list-style-type: none"> - Recurso: qualquer objetivo pode ser um recurso, na Web: uma página, uma parte de uma página, etc. - Propriedades ou atributos: aspecto específico que caracteriza o recurso. - Declarações ou valores: é o valor da propriedade. - Modelos de recipientes: <ul style="list-style-type: none"> - Bolsa: para declarar que uma propriedade tem valores múltiplos e que não há nenhuma significação à ordem na qual

<p>os valores são determinados; poderia ser usada para dar uma lista de números de parte onde a ordem de processar as partes não importa.</p> <ul style="list-style-type: none"> - Sucessão: declarar que uma propriedade tem valores múltiplos e que a ordem dos valores é significante. Por exemplo, sucessão poderia ser usada para preservar uma ordem alfabética de valores. - Alternativa: uma aplicação que usa uma propriedade cujo valor é uma coleção Alternativa está atenta que pode escolher qualquer um dos artigos na lista como apropriado.
Modelos conceituais utilizados:
<p>Ferramentas tecnológicas:</p> <ul style="list-style-type: none"> - XML
<p>Cenários:</p> <ul style="list-style-type: none"> - Cenário geral do RDF <div style="text-align: center;">  <pre> graph LR s((s)) -- p --> o[o] </pre> </div> <p>- Leia-se: o é o valor de p de s, ou; s tem uma propriedade p com um valor o, ou até mesmo; o p de s é o</p>

QUADRO 6

<p>Referência: BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American, New York, may 2001. Disponível em: <http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 19 out. 2006.</p>
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - Uma nova forma de conteúdo da web que é significativo para computadores. - A maior parte do conteúdo da Web hoje é projetada para os seres humanos lerem, e não para programas de computador manipularem significativamente. - Os computadores podem habilmente analisar páginas da Web para amostragem e tratamento de rotina --- aqui um cabeçalho, há um link para uma outra página --- mas, em geral, os computadores não têm qualquer maneira confiável para processar a semântica. - Tais agentes passando por uma página da web de uma clínica saberão que a página não apenas tem palavras-chave como "tratamento, medicina, física, terapia" (como poderia ser codificado hoje), mas também que o Dr. Hartman trabalha nesta clínica às segundas, quartas e sextas-feiras e que o <i>script</i> leva em conta um intervalo de data no formato yyyy-mm-dd e retorna horários dos compromissos. - A Web Semântica não é uma Web separada, mas uma extensão da atual, na qual é dado um significado bem definido para a informação, permitindo que computadores e pessoas trabalhem em cooperação. - A propriedade essencial da World Wide Web é a sua universalidade. - O poder de um link hipertexto é que "tudo pode se ligar a tudo". - A tecnologia Web, por conseguinte, não deve discriminar entre o projeto rascunhado e o desempenho polido, entre informações acadêmicas e comerciais, ou entre culturas, línguas, meios de comunicação social, etc. - Web Semântica será tão descentralizada quanto possível. - Para a web semântica funcionar, os computadores devem ter acesso a coleções estruturadas de informações e porções de regras de inferência, que eles possam utilizar para realizar raciocínio automatizado. - Pesquisadores da Web Semântica, em contrapartida, aceitam que os paradoxos e as perguntas sem respostas são um preço que deve ser pago para alcançar versatilidade. - No início do desenvolvimento da Web, difamadores apontaram que ela nunca seria uma biblioteca bem organizada; sem uma base de dados central e estrutura de árvore, jamais se teria a certeza de encontrar tudo. Eles tinham razão. - A Web Semântica permitirá que máquinas COMPREENDAM documentos semânticos e dados, não a fala e escrita humanas. - A Web Semântica, nomeando cada conceito simplesmente por um URI, permite a qualquer um expressar a invenção de novos conceitos com o mínimo esforço.
<p>Objetivos:</p> <ul style="list-style-type: none"> - A Web Semântica trará estrutura para o conteúdo significativo das páginas da Web, criando um ambiente em que agentes de software navegando de uma página para outra poderão facilmente realizar tarefas sofisticadas para os usuários. - A Web Semântica visa compensar o desequilíbrio no desenvolvimento da Web, que focou mais documentos processados por pessoas do que dados para inferências automáticas.

<ul style="list-style-type: none"> - Possibilitar um sistema único, mundial, porém descentralizado de representação do conhecimento. - Fornecer uma linguagem que expresse tanto os dados e as regras para raciocínio sobre os dados e que permita que as regras de qualquer sistema de representação do conhecimento existente sejam exportadas para a web.
<p>Estrutura:</p> <ul style="list-style-type: none"> - XML
<p>Modelos conceituais utilizados:</p> <ul style="list-style-type: none"> - RDF (Resource Description Framework): como conceito é a estrutura de descrição de recursos, possibilita o significado dos conteúdos, por meio de triplas, ou seja, recursos-atributos-valores, que formam redes de informações relacionadas, mas dentro de um mesmo banco de dados. - Ontologia: coleções de informação, que permite comparar ou combinar informações das duas ou mais bases de dados, onde dois ou mais termos podem estar sendo utilizados para significar a mesma coisa. A ontologia expande o RDF e combina suas informações.
<p>Ferramentas tecnológicas:</p> <ul style="list-style-type: none"> - XML (eXtensible Markup Language): linguagem de marcação extensível, permite que todos criem seus próprios rótulos, como anotações ocultas ou que marquem páginas da Web ou seções de texto em uma página. Programas, podem fazer uso dessas anotações de formas sofisticadas, mas o programador do programa tem que saber para o que o desenvolvedor da página utiliza cada anotação. Em resumo, a linguagem XML permite aos usuários adicionar uma estrutura arbitrária aos seus documentos, mas nada diz sobre o que as estruturas significam. - RDF: como ferramenta tecnológica, materializará as relações de um recurso (web site), seus atributos (características) e valores. - Ontologia: como ferramenta tecnológica é um documento ou arquivo que define formalmente as relações entre os termos. O tipo mais típico de ontologia para a Web tem uma taxonomia e um conjunto de regras de inferência. - Agentes: programas que utilizarão toda a informação semanticamente marcada pelas camadas anteriores para solucionar problemas dos usuários, como trocar essas informações com outros agentes da Web. - Assinaturas digitais: blocos criptografados de dados que os computadores e os agentes podem usar para verificar que as informações anexadas foram fornecidas por uma determinada fonte confiável. -
<p>Cenários:</p> <ul style="list-style-type: none"> - O aparelho de som estava à toda tocando "We Can Work It Out" dos Beatles, quando o telefone tocou. Quando Pete atendeu, <i>seu aparelho de telefone abaixou o som através do envio de uma mensagem para todos os outros dispositivos locais que tinham um controle de volume</i>. Sua irmã, Lucy, estava na linha falando do consultório médico: "Mamãe precisa consultar um especialista e, depois, tem que fazer uma série de sessões de fisioterapia. Duas vezes por semana ou coisa parecida. <i>Meu agente providenciará o encaminhamento</i>". Pete imediatamente concordou em oferecer o transporte. No consultório do médico, Lucy <i>instruiu seu agente de web semântica através do navegador do seu handheld. O agente rapidamente devolveu as informações sobre o tratamento prescrito à sua mãe a partir do agente do médico</i>, procurou em várias listas de fornecedores, e verificou se eles estavam no plano de seguro de sua mãe dentro de um raio de 20 milhas da sua casa e com <i>uma classificação de excelente ou muito bom na classificação de serviços confiáveis</i>. Começou-se então <i>tentar encontrar uma combinação entre horários de compromissos disponíveis (fornecidos pelos agentes de cada um dos fornecedores através de seus sites na internet) e as ocupadas agendas de Pete e Lucy</i>. (As partes enfatizadas indicam termos cujas semânticas, ou significados, foram definidos pelos agentes através da Web Semântica). Em poucos minutos, o agente apresentou-lhes um plano. Pete não gostou dele --- Hospital Universitário foi a única opção na cidade de sua mãe, e ele teria que dirigir de volta no meio do horário de pico. Ele pôs o seu próprio agente para refazer a pesquisa com preferências mais restritas sobre a localização e o horário. O agente de Lucy, tendo completa confiança no agente de Pete, no contexto da missão atual, automaticamente deu-lhe assistência fornecendo os certificados e atalhos para acessar os dados que já haviam sido coletados. Quase instantaneamente o novo plano foi apresentado: uma clínica muito mais próxima e horários também --- mas havia duas notas de advertência. Primeira, Pete teria que reagendar alguns de seus compromissos menos importantes. Ele os verificou quais eram --- não constituíam um problema. A segunda era algo sobre a lista da companhia de seguros não incluir este fornecedor como fisioterapeuta: "tipo de serviço e estado do plano de seguro solidamente verificado por outros meios", o agente assegurou-lhe. Lucy confirmou seu parecer favorável no mesmo momento em que Pete estava reclamando, "Poupem-me os pormenores", e estava tudo definido. (Claro, Pete não podia resistir aos pormenores e, mais tarde naquela noite o seu agente explicou como é que descobriu aquele fornecedor, embora ele não constasse da lista correta).

QUADRO 7

<p>Referência: PEIS, E et al. Análisis de la web semántica: estado actual y requisitos futuros. El Profesional de la Información, Barcelona v.12, n.5, sep./oct. 2003. Disponível em: <http://www.elprofesionaldelainformacion.com/contenidos/2003/septiembre/2.pdf>. Acesso em: 13 set. 2007.</p>
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - Na WS para a RI, o usuário pergunta a um agente de software, que realiza tarefas complexas de associação e inferência do conhecimento, devolvendo resultados precisos e contextualizados. - Ambiente aberto da Web atual que se estende a WS, dificulta a atuação dos agentes inteligentes. - As assinaturas digitais são umas das principais características da WS. Juntamente com as redes de confiança podem proporcionar

<p>mecanismos complexos de autenticação dos usuários e validação das permissões de acesso.</p> <ul style="list-style-type: none"> - O projeto da WS, em sua organização e funcionamento, tem muitas semelhanças com a organização de uma biblioteca, onde a semântica das obras (recursos) se descreve por meio de metadados (RDF, RDFS, DC, etc.), os usuários podem ser auxiliados pelos bibliotecários (agentes inteligentes), assim como pelo conhecimento e capacidade de raciocínio destes (ontologias) para localizar e recuperar a informação desejada. - A Web aberta e descentralizada pode certamente atrasar o desenvolvimento da WS. - Equilíbrio entre facilidade de uso e capacidade de expressão deve ser levado em conta, uma vez que a Web se difundiu pela simplicidade do HTML.
Objetivos:
<p>Estrutura:</p> <ul style="list-style-type: none"> - URI: permite a identificação de qualquer recurso na Web, o principal tipo de URI é a URL. - XML: metalinguagem que sustenta as demais camadas da WS. - RDF: linguagem de tags criada por meio da sintaxe do XML; e RDFSchema: é um vocabulário RDF que permite descrever recursos mediante uma orientação a objetos; DCMI (Dublin Core metadata initiative): conjunto de metadados descritos em RDF. - Ontologia: formada concretamente por uma taxonomia relacional e por um conjunto de axiomas ou regras de inferências. <ul style="list-style-type: none"> - OIL (ontology inference layer): primeira linguagem de representação de ontologias nos padrões do W3C, com parte descritiva (declaração de axiomas e regras) e outras de sistemas baseados em <i>frames</i> (taxonomias de classes e atributos). - Daml (Darpa's agent markup language)+OIL : uma linguagem que unifica as linguagens anteriores. - OWL (web ontology language): linguagem derivada da anterior, mas que duas sintaxe, uma para as máquinas e outras para as pessoas. - Agentes: entidades de software que recorre, filtra e processa informação contida na Web. Que devem ter as seguintes qualidades: <ul style="list-style-type: none"> - Comunicativo: entender as preferências, necessidades e objetivos dos usuários. - Autônomo: tomar decisões segundo nível de permissão do usuário. - Adaptável: apreender com os usuários, fontes de informação e outros agentes. - Assinatura digital: blocos de dados codificados que os computadores e agentes podem usar para verificar a veracidade da informação proporcionada por alguma fonte. Oferecendo a identidade do emissor e a integridade da informação.
Modelos conceituais utilizados:
Ferramentas tecnológicas:
Cenários:

QUADRO 8

<p>Referência: SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. The Semantic Web revisited. 2006. Disponível em: <eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisited.pdf>. Acesso em: 13 fev. 2008.</p>
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - A teoria semântica fornece uma conta de "significados" na qual a conexão lógica dos termos estabelece interoperabilidade entre sistemas. - Devido a ainda não haver produção em larga escala, mediação baseada em agentes, alguns comentadores argumentam que a Web semântica falhou em seu intento. - Aumentou a necessidade de compartilhamento da semântica e de uma Web de dados e de informações obtidos a partir dela. - ainda é muito aparente que a Web Semântica não está conosco em qualquer escala. - A Web original pegou o hipertexto e o fez funcionar numa escala global; a visão da RDF era a de proporcionar uma representação do conhecimento minimalista para a web. - Os URIs identificam recursos e por isso são centrais para o empreendimento da Web Semântica. - a Web Semântica desloca a ênfase de documentos para os dados. - A web semântica não pode existir sem normas cuidadosamente desenvolvidas e acordadas. - As ontologias que irão fornecer a semântica para a Web Semântica devem ser desenvolvidas, geridas e aprovadas pelas comunidades de prática cometida. - A questão para a construção de uma Web Semântica a partir destas convenções das áreas do conhecimento, com suas mudanças terminológicas, é saber quando necessitam revisão. - Uma preocupação citada freqüentemente sobre a Web Semântica: o custo do desenvolvimento e manutenção da ontologia. Em algumas áreas, os custos—não importam quão grandes—serão fáceis de recuperar. Por exemplo, uma ontologia será um poderoso e essencial instrumento em áreas bem-estruturadas tais como aplicações científicas. Em certas aplicações comerciais, o lucro potencial e ganho de produtividade ao utilizar especificações de um vocabulário bem-estruturado e coordenado excederão os custos de

<p>desenvolvimento e de uma ontologia e os custos marginais da manutenção.</p> <ul style="list-style-type: none"> - Podemos distinguir ontologias profundas de superficiais. - A Web Semântica a que aspiramos faz reutilização substancial das ontologias existentes e dos dados. - Os desafios: <ul style="list-style-type: none"> - Como é que nós eficazmente examinamos um número elevadíssimo de repositórios de informação descentralizada de escalas variáveis? - Como é que vamos alinhar e mapear duas ontologias? - Como é que podemos construir um navegador para a Web Semântica que efetivamente visualize e navegue no enorme grafo RDF conectado? - Como é que podemos estabelecer confiança e procedência dos conteúdos? - Os fatores críticos que levaram ao sucesso da Web serão também importantes para o sucesso do nosso empreendimento da Web Semântica. <ul style="list-style-type: none"> - Sociais. - Origens em decisões de concepção elementares e fundamentais sobre os princípios arquitetônicos da Web. - A representação do conhecimento e da engenharia de ontologias são para tentar captar os aspectos de conceituações compartilhadas. - Precisamos de análise científica para entender e prever os comportamentos que poderão resultar. - Ciência da Web.
<p>Objetivos:</p> <ul style="list-style-type: none"> - Esperamos que os desenvolvimentos, as metodologias, os desafios e as técnicas que discutimos aqui não somente façam crescer a Web Semântica, mas também contribuam para uma nova Ciência da Web—uma ciência que procure desenvolver, implementar e compreender os sistemas de informação distribuídos, sistemas de seres humanos e de máquinas, que operem numa escala mundial.
Estrutura:
Modelos conceituais utilizados:
<p>Ferramentas tecnológicas:</p> <ul style="list-style-type: none"> - O esquema RDF tornou-se uma recomendação em fevereiro de 2004. A RDFS tomou as especificações básicas da RDF e as estendeu para suportar a expressão de vocabulários estruturados. Ela forneceu uma linguagem de representação de ontologia mínima, o que a comunidade pesquisadora adotou amplamente. - OWL: a linguagem de Ontologia da Web, sua idéia principal é permitir a representação eficiente de ontologias que também sejam submissas aos procedimentos de decisão. Ela verifica se uma ontologia é logicamente consistente ou se um determinado conceito se insere no âmbito da ontologia. Usa a conexão fornecida pela RDF para permitir que as ontologias sejam distribuídas em todos os sistemas.
Cenários:

QUADRO 9

Referência: BERNERS-LEE. A Web 3.0, Semântica. HSM Management , n.64, p.122-128, set./out. 2007. Entrevista.
<p>Conceitos/características:</p> <ul style="list-style-type: none"> - Os computadores não entendem os dados da mesma forma que as pessoas. - Existe imensa quantidade de dados não utilizada, pois não estão em base de dados, chama-os de web de dados ou web profunda. - A WS vai além do uso dos links em hiperdocumentos, fornece uma rede de dados e permite a análise por parte das máquinas dos conteúdos em RDF. - A WS não é inteligência artificial, ela cuida da interoperação de dados bem definidos onde há relações bem estabelecidas. - Seu obstáculo é que os dados estão em bases desunidas e com formatos diferentes.
<p>Objetivos:</p> <ul style="list-style-type: none"> - A WS deve fazer com que os dados da Web profunda possam aparecer - Relacionar e conectar coisas. - Reutilizar dados - Conectar aplicações diferentes para que as pessoas possam, fazendo uma única pergunta, acessar e navegar todos os dados que lhe interessam.
<p>Estrutura:</p> <ul style="list-style-type: none"> - Linguagem de estrutura: RDF, lida com dados interoperacionais. - Linguagem de consulta: Sparql
Modelos conceituais utilizados:

Ferramentas tecnológicas:
Cenários:

QUADRO 10

MORATO, J. et al. Hacia una web semántica social. **El Profesional de la Información**, Barcelona, v.17, n.1, p.78-85, Jan./Feb. 2008.

Conceitos/características:

- A natureza dos documentos não é sempre textual na Web, e os principais sistemas de indexação automática são baseados em palavras.
- O extenso volume de documentos em linguagem natural expressada mediante a linguagem HTML também é uma dificuldade da Web.
- Problemas de polissemia e sinonímia da linguagem natural atrapalham a organização da informação.
- As iniciativas para responder esses problemas se produzem de forma parcial e local.
- Mas a Web alcançou uma fase de maturidade:
 - Simplicidade de ferramentas para a criação de páginas.
 - Os usuários não meros recuperadores-leitores de páginas, mas reclamam um papel mais ativo no compartilhamento de recursos.
 - Instituições como o W3C se tornaram referência no desenvolvimento dos padrões, oferecendo um '*norte*' para o rumo da Web.
- Propostas de solução para melhorar a organização da informação:
 - Processamento da linguagem natural (PLN).
 - Dotar buscadores de capacidade PLN
 - Problemas:
 - Ambigüidade terminológica da linguagem natural.
 - Falta de estruturação dos conteúdos dificulta interpretação automática
 - Em um ambiente heterogêneo as estruturas lingüísticas não são previsíveis.
 - As buscas simples, por um ou dois termos, invalida análise de contexto.
 - Custos.
 - Carência de mecânicos para permitir a interoperabilidade.
 - Estatística e mineração de dados.
 - Demora no desenvolvimento de técnicas para a compreensão de documentos estruturas por parte das máquinas.
 - Ambigüidade da linguagem.
 - Funcionam melhor em domínios menores específicos.
 - Criação de sistemas de organização do conhecimento.
 - Uso de linguagens documentárias diversas, principalmente as voltadas para domínios específicos.
- Mecanismos de busca já utilizam técnicas de PLN, como o Google.
- Propostas para a melhoria da organização da informação na Web:
 - Web social ou 2.0: criação de rede social que permite aos usuários certo controle sobre os recursos.
 - Soluções apontadas:
 - Permite que qualquer usuário possa discriminar os termos de descrição do recurso que será disponibilizado.
 - Suprime custos da construção de instrumentos terminológicos para um ambiente amplamente heterogêneo como a Web.
 - Uso das folksonomias:
 - Facilidade para entender a interpretação das *tags*
 - Maior implicação dos usuários na gestão dos recursos
 - Problemas apresentados:
 - Eliminar ambigüidade terminológica.
 - Interpretação automática correta.
 - Expandir ou restringir consultas
 - A incorporação de inferência mediante a representação da informação
 - Assegurar qualidade na descrição dos recursos

- Web semântica ou 3.0: possibilita a interpretação dos documentos por parte das máquinas, ampliando a qualidade da recuperação e da interoperabilidade entre diversos repositórios.

- Soluções propostas:

- Interoperabilidade: o intercâmbio de informações.

- Recursos semânticos: descrição de documentos em RDF e uso de ontologias

- Problemas apresentados:

- Falta de legibilidade das linguagens semânticas, como RDF e OWL.

- Poucas ferramentas amigáveis para a criação de documentos semânticos

- presença de duplicidade nos metadados e ontologias, que provoca desconfiança do usuário.

	Web 2.0	Web semântica
Origem	Evolução natural da Web	Proposta de Berners-Lee para revolucionar a Web.
Implantação	Muito alta	Escassa
Coordenação	Não existe	W3C
Foco	Pessoas	Máquinas
Primeiras menções	2003	1999
Expressão	Linguagem livre, expressado mediante folksonomias, palavras-chave denominadas <i>tags</i> , com problemas de sinonímia e polissemia.	Linguagem controlada, mediante linguagem de expressão de ontologias, sistema de organização do conhecimento e vocabulários de metadados.
Algumas Características	<ul style="list-style-type: none"> - Descrição de recursos para melhorar sua distribuição gratuita - Arquitetura de colaboração - Usabilidade alta - Um recurso é mais útil quanto mais uso tem 	<ul style="list-style-type: none"> - Utilização de uma linguagem padronizada com sintaxe uniforme e semântica não ambígua. - Interoperabilidade: troca de informação entre qualquer repositório. - Usabilidade escassa.

- A única forma de popularizar a WS seria através de uma aproximação com a Web social, para adaptar os usuários à criação e gestão de recursos semânticos.

- Criação de recursos que facilitem a localização e popularização de documentos semânticos.

Objetivos:

Estrutura:

Modelos conceituais utilizados:

Ferramentas tecnológicas:

Cenários:

Apêndice B. Teses e dissertações analisadas no trabalho

Nível: Tese
Título: Um modelo baseado em ontologias para representação da memória organizacional
Autor: ALMEIDA, M. B.
Ano: 2006
Orientador: Ricardo Rodrigues Barbosa
Linha de pesquisa: Gestão da informação e do Conhecimento
Instituição: Universidade Federal de Minas Gerais
Resumo: As organizações atuais estão inseridas em um ambiente turbulento, caracterizado pela alta competitividade, pela integração dos mercados e pela redução do ciclo de vida de produtos. Nesse ambiente, a informação e o conhecimento se tornaram essenciais para ao desempenho das empresas. Têm surgido, assim, discussões sobre a forma como as organizações podem preservar o conhecimento que detém para utilização futura, através de estruturas de retenção do conhecimento que buscam reproduzir as características da memória humana. O presente trabalho se insere nesse contexto e tem como objeto o estudo dos problemas advindos da prática de reter conhecimento no ambiente corporativo, e de manipular os mecanismos subjacentes à “memória das organizações”. A noção de memória tem sido utilizada por diversos autores, que citam o termo memória organizacional utilizando-o como uma metáfora para explicar fenômenos verificados no ciclo de vida das organizações. A presente pesquisa busca enriquecer a discussão sobre a memória das organizações. Para tal, desenvolve-se um modelo baseado em processos humanos e em processos tecnológicos para representar a memória organizacional e, em seguida, verifica-se em campo se tal modelo é válido. Do ponto de vista dos processos humanos, o modelo considera a necessidade de uma linguagem organizacional uniforme, a partir da qual o conhecimento individual se torna conhecimento organizacional, o qual pode ser preservado no ambiente corporativo. Do ponto de vista dos processos tecnológicos, o modelo considera a concepção do componente de representação do conhecimento de uma arquitetura denominada SMO-Sistema de Memória Organizacional. Tal componente corresponde a uma ontologia, estrutura que, além de representar a memória da organização, possibilita a criação e a manutenção de uma linguagem organizacional uniforme. Concebe-se uma ontologia representativa da memória organizacional, alimentada no ambiente de trabalho de uma organização real (CEMIG), no qual o conhecimento apreendido é formalizado e armazenado. Além disso, desenvolve-se um protótipo de um aplicativo que proporciona a apresentação do conhecimento preservado na memória da organização. A validação do modelo é realizada sobre a ontologia, a partir do protótipo.
Palavras-chave: não discriminadas.

Nível: Tese
Título: A Integração entre sistemas legislativos, terminologia e web semântica na organização e representação da informação legislativa
Autor: FEITOSA, A. L. G.
Ano: 2005
Orientador: Haruka Nakayama
Linha de pesquisa: Arquitetura da Informação
Instituição: Universidade de Brasília
Resumo: A pesquisa investigou as interfaces teóricas e aplicadas das áreas de Terminologia e Web Semântica, com vistas à utilização em Sistemas de Informação Legislativa. O objetivo foi identificar fundamentos teóricos e metodológicos, bem como princípios e diretrizes referentes a essas áreas do conhecimento, cuja sistematização pudesse fornecer insumos para a elaboração de uma proposta metodológica de organização, representação e recuperação da informação legislativa. Para tanto, o estudo foi realizado nas seguintes etapas: a) levantamento dos fundamentos teóricos e metodológicos referentes a conceito, sistema de conceitos, controle terminológico, linguagem documentária e web semântica, tendo em vista suas aplicações em sistemas de informação legislativa; b) levantamento e análise das características de 29 Sistemas de Informação Legislativa brasileiros, compreendendo o Senado Federal, a Câmara dos Deputados, a Câmara Legislativa do Distrito Federal e as Assembleias Legislativas Estaduais; c) levantamento e análise das características dos sistemas NUJR, SAPL, NJUR-DF e LEGIN; d) elaboração de proposta metodológica para a organização, representação e recuperação de informações em sistemas legislativos. A proposta elaborada fundamenta-se: a) no losango semiótico, uma adaptação do modelo de triângulo semiótico de Ogden e Richards (1972), em face dos fundamentos teóricos e metodológicos da Terminologia e da Web Semântica; b) no Formato de Marcação de Terminologias (FORMATER), elaborado nesta pesquisa; c) no Formato de Marcação Legislativa (FORMALEGIS), elaborado nesta pesquisa. Conclui-se que a proposta metodológica apresentada é aplicável a Sistemas de Informação Legislativa e, feitas as adaptações necessárias, passível de extensão a outros sistemas de informação.
Palavras-chave: Informação legislativa. Sistema legislativo. Web semântica. Terminologia. Linguagem documentária. Indexação. Ontologia. XML. OWL. RDF.

Nível: Tese
Título: Autoria de documentos para a web semântica: um ambiente de produção de conhecimento baseado em ontologias
Autor: OLIVEIRA, E. C.
Ano: 2006
Orientador: Mamede Lima-Marques
Linha de pesquisa: não discriminada
Instituição: Universidade de Brasília
Resumo: Nesta tese, apresentamos o momento da autoria de documentos como situação potencial para a representação do conhecimento e a conseqüente publicação de documentos. Sistemas baseados em ontologia para a representação da informação e autoria na Web Semântica estão surgindo e provendo soluções como anotação semântica, extração da informação, marcação ontológica, etc. Ao analisar esses sistemas, descobrimos as características que eles deveriam ter, a fim de que produzam e suportem documentos em conformidade com as recomendações da Web Semântica e conseqüentemente usufruam de suas potencialidades. Apresentamos aqui uma análise de ferramentas baseadas em ontologias, a partir de um critério de análise específico desenvolvido para identificar as características, requisitos e conformidade com as recomendações existentes, identificadas na revisão de literatura. O resultado da análise revelou que nesse novo paradigma de produção, os autores deverão desempenhar tarefas de estruturação do conhecimento com os quais eles ainda não estão familiarizados. Identificamos a necessidade de abordar essas questões no âmbito da Ciência da Informação, buscando dar apoio a autores no desenvolvimento de conteúdos bem estruturados e representados, a partir do potencial que têm as atuais tecnologias baseadas em ontologias. Após tratar dessas questões, propomos com esse trabalho um conjunto de requisitos e um modelo de arquitetura de ambiente de autoria baseado em ontologia para a Web Semântica, composto por módulos que sugerem novas funções e integram tecnologias correntes, fazendo uso de possibilidades existentes e propondo o desenvolvimento de novos requisitos.
Palavras-chave: Produção de documentos. Web Semântica. Ontologia. Arquitetura da informação. Representação do conhecimento. Ambiente de autoria.

Nível: Tese
Título: Recuperação da informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação.
Autor: FERNEDA, E.
Ano: 2003
Orientador: Johanna Wilhelmina Smit
Linha de pesquisa: Análise Documentária
Instituição: Universidade de São Paulo
Resumo: Desde o seu nascimento, a Ciência da Informação vem estudando métodos para o tratamento automático da informação. Esta pesquisa centrou-se na Recuperação de Informação, área que envolve a aplicação de métodos computacionais no tratamento e recuperação da informação, para avaliar em que medida a Ciência da Computação contribui para o avanço da Ciência da Informação. Inicialmente a Recuperação de Informação é contextualizada no corpo interdisciplinar da Ciência da Informação e são apresentados os elementos básicos do processo de recuperação de informação. Os modelos computacionais de recuperação de informação são analisados a partir da categorização em "quantitativos" e "dinâmicos". Algumas técnicas de processamento da linguagem natural utilizadas na recuperação de informação são igualmente discutidas. No contexto atual da Web são apresentadas as técnicas de representação e recuperação da informação desde os mecanismos de busca até a Web Semântica. Conclui-se que, apesar da inquestionável importância dos métodos e técnicas computacionais no tratamento da informação, estas se configuram apenas como ferramentas auxiliares, pois utilizam uma conceituação de "informação" extremamente restrita em relação àquela utilizada pela Ciência da Informação.
Palavras-chave: Informação. Ciência da Informação. Ciência da Computação. Recuperação de informação. Modelos de recuperação de informação.

Nível: Dissertação
Título: Representação de informação e conhecimento: estudo das diferentes abordagens entra a Ciência da Informação e a Ciência da Computação
Autor: FURGERI, S.
Ano: 2006
Orientador: Raimundo Nonato Macedo dos Santos
Linha de pesquisa: Produção e Disseminação da Informação
Instituição: Pontifícia Universidade Católica de Campinas
Resumo: A Ciência da Informação vem estudando formas de representação da informação e do conhecimento visando à recuperação

da informação. Esta pesquisa tem seu foco na representação do conhecimento e da informação, procurando investigar quais são os pontos convergentes e divergentes entre as linguagens documentárias da Ciência da Informação e as linguagens de marcação desenvolvidas e utilizadas na Ciência da Computação, tendo em vista identificar ações, teorias e processos necessários para uma maior integração entre as duas áreas. Para isso, faz-se uma revisão dos elementos fundamentais necessários à representação da informação e do conhecimento no âmbito da Ciência da Informação. Para tornar possível a comparação entre áreas, apresentam-se os modelos mais consagrados de representação do conhecimento e da informação provenientes da Ciência da Informação, tais como metadados, tesouros e ontologias. No âmbito da Internet, apresentam-se as técnicas de representação com o uso das linguagens de marcação mais utilizadas e suas contribuições para o desenvolvimento da Web Semântica. Encerra-se apresentando uma proposta de estrutura de representação para recursos informacionais, especialmente os disponibilizados pela Internet. A proposta foi desenvolvida a partir dos recursos existentes na Ciência da Computação, particularmente, os oferecidos pela linguagem XML. Contempla a definição de uma ontologia e culmina com a criação de uma estrutura em XML para armazenar metadados de artigos eletrônicos.
Palavras-chave: Metadados. Linguagens de marcação. Ciência da Informação. Ciência da Computação. Modelos de representação. Ontologias.

Nível: Dissertação
Título: Ontologias de domínio: estudo das relações conceituais e sua aplicação
Autor: SALES, L. F.
Ano: 2006
Orientador: Maria Luiza de Almeida Campos; Co-orientador: Hagar Espanha Gomes
Linha de pesquisa: Representação, Gestão e Tecnologia da Informação
Instituição: Universidade Federal Fluminense
Resumo: No âmbito da elaboração de Ontologias, uma problemática que tem se colocado é a ausência de um padrão teórico-metodológico para sua elaboração, o que implica na criação a criação de regras de inferências consistentes que possam ser interpretáveis pela máquina. Propõe-se então uma sistematização das relações apresentadas nas literaturas da Ciência da Informação, da Terminologia, da Ciência da Computação e da Bioinformática. Na Ciência da Informação, as Categorias Fundamentais proporcionam um modelo de representação de um domínio, mas não explicitam a relação entre elas. A Ciência da Computação explicita as relações, mas não fornece um contexto. A sistematização realizada busca unir as duplas de categorias (relações categoriais) com as relações potenciais entre elas (relações formais). A partir da seleção de um corpus da Gene Ontology, foi feita uma análise das definições, a fim de identificar as relações já mencionadas na literatura ou descobrir a existência de novas relações. Apesar da Gene Ontology apresentar padrões para as definições, nem sempre foi possível identificar facilmente as relações. Conclui-se que relações diferentes podem surgir em diferentes domínios e que definições sistematizadas são indispensáveis para o estabelecimento seguro de relações formais.
Palavras-chave: Ontologias. Relações conceituais. Recuperação de informação.

Nível: Dissertação
Título: Ontologias no suporte a portais semânticos
Autor: JORGE, M. A.
Ano: 2005
Orientador: Marcello Peixoto Bax
Linha de pesquisa: Gestão da Informação e do Conhecimento
Instituição: Universidade Federal de Minas Gerais
Resumo: Portais semânticos aparecem como uma evolução natural dos tradicionais portais web e podem ser entendidos como portais de informações que utilizam os padrões de representação de informação propostos pelo W3C para a chamada Web Semântica. As ontologias são a base desses novos portais web, propiciando melhor organização das informações, além de outras vantagens. O objetivo deste estudo é compreender como o uso de ontologias promove a organização da informação e agrega valor a esses portais. Para atingir este objetivo, o trabalho investiga as principais ferramentas semânticas baseadas em ontologias existentes na literatura científica e elabora uma ontologia que serve de sustentação para a proposição de um portal semântico a ser utilizado por uma comunidade universitária. Após uma pré-especificação do portal semântico, sua arquitetura funcional é sugerida e as funções semânticas são apresentadas e detalhadas através do uso da ontologia construída anteriormente. As ferramentas semânticas investigadas são utilizadas no contexto de especificação do portal semântico proposto e exemplos ilustrativos, baseados na ontologia elaborada, são extraídos. Com o trabalho espera-se contribuir para a discussão acerca dos benefícios que uma ontologia pode trazer para a organização de conceitos, promovendo a reutilização e o compartilhamento de informações em um determinado domínio do conhecimento.
Palavras-chave: não discriminada

Nível: Dissertação
Título: Tesouros e ontologias sob a luz da teoria comunicativa da terminologia
Autor: SALES, R.
Ano: 2008
Orientador: Lígia Maria Arruda Café
Linha de pesquisa: Fluxos de Informação
Instituição: Universidade Federal de Santa Catarina
Resumo: O acelerado desenvolvimento das tecnologias de informação e comunicação promove um avanço vertiginoso na veiculação do conhecimento. Tal fato incide diretamente na comunicação entre especialistas que, por meio de linguagens de especialidade, laçam mão de terminologias próprias e específicas para transmitirem conteúdos informacionais das mais diversas áreas científicas e tecnológicas. Linguagens Documentárias são modelos de representação do conhecimento que, servindo como instrumentos de controle terminológico, auxiliam o processo de indexação e recuperação de documentos por assunto. O tesouro é uma linguagem documentária caracterizada pela especificidade e pela complexidade existente no relacionamento entre os termos que comunicam o conhecimento especializado. A ontologia é um modelo de representação do conhecimento que, a exemplo do tesouro, é utilizada para representar e recuperar informação por meio de uma estrutura conceitual (no caso da ontologia o meio de ação é o informático). A presente pesquisa consiste sobretudo na identificação, por meio de uma pesquisa documental, de aspectos que aproximam e distanciam os fundamentos do tesouro e o da ontologia. Para tanto, foi empregado, no plano metodológico, o Método de Análise de Conteúdo de Laurence Bardin. No plano teórico, a pesquisa está fundamentada pela Teoria Comunicativa da Terminologia (TCT) de Maria Teresa Cabré, que zela pelo caráter comunicativo no ambiente científico e técnico, baseando-se no uso social da linguagem efetivamente utilizada por especialistas. Para a constituição do corpus de análise foram utilizadas as bases de dados da Library and Information Science Abstracts (LISA), da Wilson Library Literature and Information Science Full Text e da Biblioteca Digital de Teses e Dissertações do IBICT (BDTD). A análise dos resultados e interpretações proporcionaram significativas contribuições aos estudos dos modelos de representação do conhecimento, visto que inúmeras características semelhantes e diferentes entre tesouros e ontologias foram identificadas e comentadas. Constatou-se que ambos os instrumentos se aproximam no que diz respeito ao esclarecimento dos aspectos relativos aos termos e suas estruturas conceituais. Suas diferenças se acentuam na esfera das aplicações, pois os recursos informáticos que suportam as ontologias as concedem objetivos que vão além daqueles atribuídos aos tesouros.
Palavras-chave: Tesouro. Ontologia. Linguagem documentária. Teoria comunicativa da terminologia. Método de análise de conteúdo

Nível: Dissertação
Título: Inserção da biblioteca digital de teses e dissertações no contexto da web semântica: construção e uso da ontologia
Autor: MOLOSSI, S.
Ano: 2008
Orientador: Lígia Maria Arruda Café
Linha de pesquisa: Fluxos de Informação
Instituição: Universidade Federal de Santa Catarina
Resumo: A problemática da explosão informacional detectada em meados dos anos 40 aliada as constatações das deficiências dos sistemas convencionais de consulta motivaram a aplicação das novas tecnologias da informação e comunicação no desenvolvimento de modernos modelos e padrões de representação e recuperação da informação. Dentre eles, encontra-se a Web Semântica que associa significado à informação, permitindo o trabalho cooperativo entre computadores e pessoas. Com o intuito de contribuir para o desenvolvimento desta área no contexto brasileiro, esta pesquisa tem por objetivo analisar as tecnologias de recuperação da informação sob a perspectiva da Web Semântica para a adaptação e aperfeiçoamento do método de recuperação atualmente adotado pela Biblioteca de Teses e Dissertações (BDTD) do Instituto Brasileiro de Informação e Tecnologia (IBICT). Do ponto de vista metodológico, este estudo analisa a estrutura da BDTD, bem como os metadados <i>Título</i> , <i>Assunto</i> e <i>Resumo</i> , a fim de extrair os termos semanticamente relevantes desses campos e desenvolver uma ontologia para um domínio específico. Para o desenvolvimento da ontologia foram utilizados a ontoKEM como ferramenta de apoio a elaboração do projeto e a documentação do trabalho e o <i>Protégé</i> para sua construção. Foi ainda adotado o arquivo OWL como meio de integrar essas duas visões no desenvolvimento (o projeto e a construção). Os resultados obtidos indicam que estudos de ontologia conduzem à interoperabilidade entre os padrões existentes e essa nova proposta contextualizada pela Web Semântica e suas tecnologias. No que diz respeito ao ambiente da BDTD, foi identificado que além da ontologia desenvolvida e de uma programação nos seus agentes de busca, a marcação semântica é peça importante nesse cenário. Como conclusão, são propostas melhorias para as técnicas de representação e recuperação da informação, utilizadas pela BDTD, respeitando as peculiaridades de sua estrutura. Por fim, sugere-se como trabalhos futuros a realização de um estudo posterior acerca da possibilidade dos metadados atenderem a função de marcação semântica.
Palavras-chave: Web Semântica. Metadados. Ontologia. Recuperação da informação. BDTD. TEDE. IBICT.

Nível: Dissertação
Título: Metodologia para validação de ontologias: o caso ORBIS_MC

Autor: BEDIN, S. P. M.
Ano: 2007
Orientador: Lígia Maria Arruda Café
Linha de pesquisa: Fluxos de Informação
Instituição: Universidade Federal de Santa Catarina
<p>Resumo: O desenvolvimento científico e tecnológico vem trazendo em seu escopo novas estruturas de conteúdos. Relações de toda ordem montam teias que ligam dados, informações, pessoas e locais em um mesmo ambiente não mais físico. A velocidade determinante deste novo século tem gerado inquietações acerca da nova situação que envolve todas as áreas do conhecimento. Grandes volumes de conteúdos precisam ser recuperados com rapidez e precisão atendendo às expectativas dos usuários. Sistemas de Recuperação de Informação são desenvolvidos na tentativa de solucionar esta questão. A pesquisa se insere no universo das linguagens aplicadas nesses sistemas e sua motivação é fruto da leitura e reflexão de estudos anteriores sobre o tema, envolvendo aspectos da sua estrutura, conceitos, relacionamentos e variações. Abordando a linguagem natural e suas novas formas de representação, são estudadas as ontologias de um Sistema de Recuperação de Informações que utiliza a tecnologia KMAI^r. Esta tecnologia adota técnicas de inteligência artificial e ontologias como forma de recuperação da informação em documentos não-estruturados disponíveis em mídias escritas: jornais, revistas e sites eletrônicos com código aberto. O sistema-alvo de estudo, desenvolvido para o Observatório Regional Base de Indicadores de Sustentabilidade Metropolitano de Curitiba- ORBIS_MC, busca estruturar informações que vão subsidiar políticas públicas de desenvolvimento para a região metropolitana por meio de monitoramento, recuperação e composição da base do conhecimento sobre Saúde Ambiental, Saúde Humana e Violência Urbana na Região Metropolitana de Curitiba. O sistema conta com um dicionário de termos construído por especialistas nas áreas definidas como objeto de representação. A presente pesquisa propõe uma metodologia de validação das ontologias que integram este dicionário, considerando os critérios lingüísticos que reflitam o uso da língua no contexto comunicativo, visto que se trata de uma linguagem utilizada para recuperar conteúdos cuja função é comunicar. Apoiando-se no estudo de publicações já reconhecidas nas áreas de Ciência da Informação e Lingüística, foram sistematizadas recomendações sobre o estabelecimento das relações semânticas entre os termos, a fim de definir critérios de validação de ontologias sob o ponto de vista lingüístico. Esses critérios foram aplicados em um “corpus” de 28 termos e 163 relações semânticas. O resultado da análise apresenta 82,21% de relações semânticas consideradas consistentes e 17,79% inadequadas. Para este último conjunto, foram tecidas recomendações. A aplicação de uma metodologia de fundamentação lingüística sob a perspectiva do uso foi a obtenção de uma ontologia atualizada, tanto no seu aspecto lexical como semântico, aumentando a eficiência de recuperação da informação.</p>
Palavras-chave: Ontologias. Validação de ontologias. Recuperação da informação.

Nível: Dissertação
Título: Recomendações para construção de taxonomia em portais corporativos
Autor: VITAL, L. P.
Ano: 2007
Orientador: Lígia Maria Arruda Café
Linha de pesquisa: Fluxos de Informação
Instituição: Universidade Federal de Santa Catarina
<p>Resumo: A gestão do conhecimento nas organizações, processo que envolve a geração, a codificação, coordenação e transferência do conhecimento, é imprescindível diante da competitividade da sociedade atual. No nível organizacional, este processo utiliza diferentes ferramentas, dentre elas os portais corporativos, que surgem como uma forma de unificar informações por um único ponto de acesso, compartilhando conhecimentos e possibilitando seu desenvolvimento. Entretanto, para que um portal corporativo seja eficiente é necessária a criação de meios de acesso, busca e a publicação da informação por qualquer usuário cadastrado no portal. Neste contexto de organização e recuperação da informação em ambientes corporativos, surgem os modelos de representação do conhecimento como instrumentos de apoio a estes processos. Entre eles está a taxonomia, foco central desta pesquisa. Para que uma taxonomia venha a ser utilizada plenamente, é importante que sua construção siga métodos precisos e objetivos. No entanto, as metodologias atualmente disponíveis não apresentam orientação suficientemente detalhada, o que deixa lacunas para entendimentos indevidos na sua construção e aplicação. Diante deste quadro, analisamos nesta pesquisa as metodologias de elaboração de taxonomias disponíveis na literatura com o objetivo de propor recomendações na sua composição. Desta forma, a presente dissertação examinou os modelos de desenvolvimento de taxonomias encontrados na literatura da área de Ciência da Informação entre os anos de 2000 e 2005, objetivando identificar e propor diretrizes para sua construção, tendo como referencial a funcionalidade relacionada à categorização, identificada na literatura da área de portais corporativos. A pesquisa se configura como um estudo exploratório qualitativo. Analisamos os modelos sob a ótica da funcionalidade categorização nos aspectos de: 1) Construção da taxonomia (representação do conhecimento), nos sub-itens: estabelecimento das categorias gerais, coleta dos termos, análise dos termos selecionados, controle da diversidade de significação e construção dos relacionamentos semânticos; e 2) Aplicação da taxonomia, nos sub-itens: representação da informação (indexação ou classificação), organização da informação recuperada. Esses aspectos de análise foram definidos com base na literatura da área de Ciência da Informação e de portais corporativos. A análise resultou em recomendações para o aprimoramento de metodologias para construção de taxonomias. Sendo assim, foi possível apresentar proposições embasadas nas áreas de Classificação, Indexação, Lingüística e Documentação, cujas</p>

discussões teóricas e metodológicas auxiliam o processo de construção de uma taxonomia.
Palavras-chave: Organização da informação. Representação do conhecimento. Modelo de representação. Portal corporativo. Taxonomia.

Nível: Dissertação
Título: Recuperação de informações sobre <i>log</i> de eventos apoiada em ontologia
Autor: ESBÍZARO, A. L. D.
Ano: 2006
Orientador: Marisa Bräscher Basílio Medeiros
Linha de pesquisa: Arquitetura da Informação
Instituição: Universidade de Brasília
Resumo: A pesquisa propõe o desenvolvimento de uma ontologia de log de eventos do sistema operacional Microsoft Windows Server 2003, visando facilitar a recuperação de informações para apoiar o entendimento e a resolução de logs de eventos. Os sistemas de solução de eventos apresentam problemas relativos à falta de instrumentos para acesso ao conteúdo semântico dos registros e para armazenamento do conhecimento adquirido sobre log de eventos e à ausência de compreensão comum sobre logs de eventos e de relacionamentos entre os diferentes fatores que levam à ocorrência de erros nos sistemas operacionais. As etapas da metodologia abrangem: análise das características do site oficial <i>Events and Errors Message Center</i> e o programa aplicativo Event Viewer (Visualizar Eventos), levantamento de informações sobre log de eventos, modelagem da ontologia, simulação do uso da ontologia numa interface de busca sobre log de eventos e análise dos resultados obtidos na busca usando as ferramentas da Microsoft e na busca apoiada por ontologia. O resultado da pesquisa revela dificuldades no entendimento dos log de eventos, devido à dispersão das informações sobre logs nas ferramentas analisadas, à falta de relacionamento entre conceitos envolvidos na resolução dos logs, à ausência total de informações sobre determinados logs identificados. De maneira geral, conclui que a ontologia auxilia no entendimento dos log de eventos, mas que sua resolução depende de um nível mais profundo de descrição ontológica.
Palavras-chave: Log de eventos. Ontologia. Recuperação de informação.

Nível: Dissertação
Título: Ontologias: indexação e recuperação de fotografias baseadas na técnica fotográfica e no conteúdo da imagem
Autor: MIRANDA, A. S. S.
Ano: 2007
Orientador: Orientador: Mirian Paula Manini; Co-orientador: Mamede Lima-Marques
Linha de pesquisa: Gestão da Informação e do Conhecimento
Instituição: Universidade de Brasília
Resumo: Este trabalho aborda as técnicas de representação da imagem por conceito com o objetivo de fundamentar a construção de uma ontologia de descrição de fotografias através da utilização de padrões da Web Semântica, como as linguagens OWL e RDF Schema. A ontologia de descrição consiste numa representação ontológica do conteúdo visual da imagem, além de definir um esquema de descrição de imagens. Para validar a ontologia de descrição, foi desenvolvido, como parte da metodologia desta pesquisa, um sistema de descrição e recuperação de imagens que usa a ontologia internamente para descrever e recuperar fotografias. A idéia deste trabalho é enriquecer a descrição de imagens com a associação de conceitos e instâncias extraídos da ontologia. A ontologia de descrição foi construída levando em consideração a informação presente na imagem e dados oriundos da técnica fotográfica usada no momento da produção da fotografia. O resultado da descrição da imagem usando o sistema protótipo de descrição é um metadado em RDF (Resource Description Framework) que descreve o conteúdo visual da imagem, somado com a descrição da técnica fotográfica empregada na produção da fotografia.
Palavras-chave: Ontologia. Indexação de imagem.

Nível: Dissertação
Título: Um arcabouço teórico para autoria de documentos visando atenuar o surgimento do fenômeno da ambigüidade
Autor: SANTOS, D. P. L. dos
Ano: 2006
Orientador: Mamede Lima-Marques
Linha de pesquisa: Arquitetura da Informação
Instituição: Universidade de Brasília
Resumo: As ambigüidades e seus efeitos na recuperação e interpretação da informação constituem-se no problema tratado por esta dissertação. Este trabalho aborda o problema no momento da autoria dos documentos, ao invés da abordagem convencional dos

<p>sistemas de Processamento Automático de Linguagem Natural, que lidam com as ambigüidades a posteriori. Parte-se do pressuposto que o autor do documento sabe exatamente o que quer comunicar e que, portanto, o mesmo está apto a auxiliar o receptor no processo de comunicação, sanando ou minimizando a ocorrência de ambigüidades na recuperação e interpretação da informação. Assim, o principal objetivo do trabalho é criar um arcabouço teórico que permita alçar o autor do documento a uma posição mais participativa no processo de comunicação, auxiliando o leitor através de informações de cunho morfológico, sintático e semântico sobre o conteúdo do documento. Para tal, a obra de Bräscher relativa ao tratamento automático de ambigüidades na recuperação da informação, bem como os ambientes de autoria baseados em ontologias, de Oliveira, surgem como principais ferramentas teóricas para viabilizar o arcabouço proposto. Ao final, constata-se que o arcabouço teórico mostra-se eficaz para atenuar o surgimento de ambigüidades na recuperação e interpretação da informação, principalmente no campo léxico-morfológico, e também nas esferas sintática e predicativa. Entretanto, percebeu-se que o arcabouço mostrou-se ineficaz com relação às ambigüidades de cunho pragmático e semântico.</p>
<p>Palavras-chave: Ambigüidade. Ontologia. Gramática de valências. Ambiente de Autoria.</p>

Nível: Dissertação
Título: Ambigüidades da língua portuguesa: recorte classificatório para a elaboração de um modelo ontológico
Autor: SILVA, L. B. da
Ano: 2006
Orientador: Mamede Lima-Marques
Linha de pesquisa: Arquitetura da Informação
Instituição: Universidade de Brasília
<p>Resumo: Ambigüidade é um fenômeno lingüístico que permite que uma expressão possa ter várias interpretações válidas. Tal fenômeno pode apresentar-se de muitas formas, tem o poder de mudar o entendimento de um texto convertendo completamente seu sentido. Com tamanho poder, a ambigüidade causa problemas em textos cujo objetivo é informar e esclarecer. Para resolver uma ambigüidade é necessário conhecê-la de forma horizontal e vertical. Saber qual a largura e profundidade do problema é fundamental para propor novas soluções. Esta pesquisa buscou desenvolver um mapeamento completo das ambigüidades da Língua Portuguesa. Durante o seu desenvolvimento foi possível identificar: (i) os tipos de ambigüidades presentes na Língua Portuguesa - ressalta-se que as ambigüidades pesquisadas se referem apenas ao contexto brasileiro; (ii) seus tipos e subtipos; (iii) as ambigüidades que devem ser resolvidas ou preservadas; (iv) por quantos nomes uma mesma ambigüidade é conhecida. Uma base de testes foi desenvolvida com aproximadamente sessenta frases ambíguas para vários tipos de ambigüidades. Para a definição do modelo foi utilizado o Protégé e a linguagem de ontologias da Web. Com o uso dessas ferramentas foi possível desenvolver uma ontologia onde é implementada parte do conhecimento produzido por esta pesquisa.</p>
<p>Palavras-chave: Ambigüidade. Taxonomia. Ontologia. Terminologia. Web Semântica. Linguagens de ontologias.</p>

Nível: Dissertação
Título: Web semântica: uma análise focada no uso de metadados
Autor: ALVES, R. C. V.
Ano: 2005
Orientador: Plácida L. V. A. da Costa Santos
Linha de pesquisa: Informação e Tecnologia
Instituição: Universidade Estadual Paulista
<p>Resumo: Atualmente a nossa sociedade, denominada sociedade da informação, vem sendo caracterizada pela valorização da informação, pelo uso cada vez maior de tecnologias de informação e comunicação e pelo crescimento exponencial dos recursos informacionais disponibilizados em diversos ambientes, principalmente na Web. Essa realidade trouxe algumas mudanças no acesso automatizado às informações. Se por um lado temos uma grande quantidade de recursos informacionais disponibilizados, por outro temos como consequência problemas relacionados à busca, localização, acesso e recuperação dessas informações em ambientes digitais. Nesse contexto, o problema que originou essa pesquisa está relacionado com a dificuldade na busca e na recuperação de recursos informacionais digitais na Web e a ausência de tratamento adequado para a representação informacional desses recursos. O maior desafio para a comunidade científica no momento está na identificação de padrões e métodos de representação da informação, ou seja, na construção de formas de representação do recurso informacional de maneira a proporcionar sua busca e recuperação de modo mais eficiente. Assim, a proposição apontada nesse trabalho como solução do problema refere-se ao estabelecimento da Web Semântica e a aplicação de padrões de metadados para a representação da informação, pois são consideradas como iniciativas importantes para proporcionar uma melhor estruturação e representação dos recursos informacionais em ambientes digitais. Com uma metodologia baseada na análise exploratória e descritiva do tema a partir da literatura disponível, apresenta-se uma análise da Web Semântica como uma nova proposta para organização dos recursos informacionais na Web e as ferramentas tecnológicas que permeiam sua construção, com enfoque no uso de metadados como elemento fundamental para proporcionar uma melhor representação dos recursos informacionais disponibilizados na Web e sua posterior recuperação. A proposta da Web Semântica é disponibilizar recursos informacionais melhor estruturados e representados, formando uma rede de informações conectadas que por</p>

<p>meio de ferramentas tecnológicas, tais como: os agentes de software, a linguagem de marcação XML, arquitetura de metadados RDF, ontologias e, principalmente, padrões ou formatos e metadados. Como resultado pode-se destacar que a implementação da Web Semântica requer o trabalho conjunto das várias ferramentas tecnológicas estudadas e que proporcionará em pequena, média e grande escala a tão necessária estruturação e representação informacional dos recursos e conseqüentemente sua melhor recuperação. Além disso, foi possível verificar que as tecnologias da Web Semântica convergem para a área de Ciência da Informação, estabelecendo uma estreita relação na questão da representação do conhecimento, principalmente com relação ao uso de metadados que são considerados essenciais para se estabelecer uma boa representação dos recursos informacionais na rede. Sabendo que a representação da informação é necessária em qualquer ambiente para proporcionar uma recuperação mais eficiente, podemos considerar os metadados como ferramentas essenciais para estabelecer a representação dos recursos informacionais no ambiente da Web Semântica como instrumentos para a construção de uma rede de conhecimentos e recuperação da informação de modo mais eficiente.</p>
<p>Palavras-chave: Web Semântica. Metadados. Representação da informação. Web. Ontologias. Arquitetura de metadados RDF. Linguagem de marcação XML.</p>

Nível: Dissertação
Título: Padrões de representação e descrição de recursos informacionais em bibliotecas digitais na perspectiva da Ciência da Informação: uma abordagem do MarcOnt Initiative na era da Web Semântica
Autor: CASTRO, F. F. de
Ano: 2008
Orientador: Plácida L. V. A. da Costa Santos
Linha de pesquisa: Informação e Tecnologia
Instituição: Universidade Estadual Paulista
<p>Resumo: A relevância tanto da Web quanto das bibliotecas digitais para os diversos ramos da ciência, tem impulsionado pesquisadores e comunidades científicas a buscar soluções de integração, intercâmbio e entendimento semântico sobre os conteúdos que nelas circulam. Verifica-se, no cenário atual, uma forte tendência para a disponibilização de conteúdos digitais e a falta de uso de padrões adequados para a representação e a descrição desses recursos em ambientes digitais, o que é apontado pela literatura como uma grande preocupação para a comunidade científica e focada neste trabalho enquanto problema da pesquisa. Nesse contexto, a proposição desta pesquisa é verificar na literatura científica a aplicabilidade e a funcionalidade das tecnologias e instrumentos disponíveis como padrões de representação e descrição de conteúdo e forma, no âmbito das bibliotecas digitais, com o objetivo de identificar e de caracterizar o uso e a aplicação do MarcOnt enquanto ferramenta para a construção de formas de representação da informação para bibliotecas digitais na atualidade, destacando-o como instrumento para descrição dos aspectos semânticos dos conteúdos digitais nestas ambiências. Como metodologia, o presente trabalho caracteriza-se por ser uma pesquisa de análise exploratória e descritiva do tema, com o qual é possível obter subsídios para um maior entendimento e compreensão mais detalhada do MarcOnt e as tecnologias que estão abarcadas no âmbito das bibliotecas digitais para a construção Web Semântica. Como resultado, procurou-se demonstrar a aplicação da ferramenta MarcOnt em um ambiente informacional específico, a Biblioteca Digital Semântica Jeromedl, para analisar nos fundamentos da Catalogação as formas de representação dos recursos informacionais. O que nos levou a definir categorias para análise futura do sistema em comparação com o padrão de metadados MARC21. Além disso, foi possível demonstrar a existência de um ambiente colaborativo, o Portal MarcOnt, o qual se apresenta como um espaço para a construção de descrições bibliográficas semânticas. Assim, o padrão de descrição bibliográfica semântica MarcOnt, oferece a possibilidade de garantir uma melhora na performance de interoperabilidade, não somente entre padrões de metadados, mas também entre sistemas e ambientes informacionais digitais.</p>
<p>Palavras-chave: MarcOnt. MARC21. Web Semântica. Bibliotecas digitais. Representação da informação. Catalogação automatizada. Padrões de metadados. Informação e tecnologia.</p>

Nível: Dissertação
Título: Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade
Autor: GRACIO, J. C. A.
Ano: 2002
Orientador: Plácida L. V. A. da Costa Santos
Linha de pesquisa: Informação e Tecnologia
Instituição: Universidade Estadual Paulista
<p>Resumo: Os avanços na Ciência e Tecnologia, particularmente na informática e nas telecomunicações, fizeram com que a Internet se tornasse uma das mais importantes fontes de informação. Ao mesmo tempo, a carência de dados adequados sobre os recursos disponíveis na Web leva à necessidade de criação de elementos descritivos que possam contribuir para a representação de tais recursos. Nesse sentido, a utilização de metadados parece ser a melhor solução para o problema. Metadados podem ser definidos como um conjunto de elementos que descrevem as informações contidas em um recurso, permitindo, assim, sua busca e recuperação. Desse modo, tem-se os metadados em, seu conceito, funcionamento, padrões existentes e formas de utilização, bem como as</p>

características e a estrutura do padrão Dublin Core para descrição de recursos na Web e, ainda, um estudo de diversos projetos que utilizam metadados. A análise dos projetos indica que o padrão Dublin Core, composto de um conjunto de 15 elementos, vem sendo amplamente utilizado por sua simplicidade de descrição, extensibilidade e interoperabilidade. Essa última característica é um requisito importante, considerando-se os diversos padrões de metadados existentes, pois a correspondência entre eles é fundamental para a troca de informações. Os resultados obtidos indicam que estudos de metadados conduzem à interoperabilidade entre os padrões existentes e sua representação por linguagens de marca, como a linguagem XML, juntamente com a arquitetura RDF.
Palavras-chave: Metadados. Dublin Core. Interoperabilidade. Descrição de recursos na Internet. Busca e recuperação de informação na Web.

Nível: Dissertação
Título: A construção do conhecimento científico: a Web Semântica como objeto de estudo
Autor: PINHEIRO, C. B. F.
Ano: 2008
Orientador: Edberto Ferneda
Linha de pesquisa: Informação e Tecnologia
Instituição: Universidade Estadual Paulista
Resumo: A Sociologia da Ciência ou os Estudos Sociais da Ciência têm procurado analisar a estrutura das disciplinas científicas em relação às suas práticas sociais, procurando oferecer novas perspectivas sobre a construção do saber e o desenvolvimento científico e tecnológico. Um conceito importante para o desenvolvimento deste trabalho é o de comunidade científica caracterizada pela prática de uma especialidade, por uma formação teórica comum, pela circulação abundante de informação no interior do grupo. Este estudo tem por objetivo caracterizar a pesquisa brasileira em Web Semântica, considerado um tema na fronteira de pesquisa de duas matrizes disciplinares: a Ciência da Computação e a Ciência da Informação. Optou-se por um estudo cienciométrico, tendo como fonte para coleta dos dados o Curriculum Lattes de pesquisadores doutores e nas dissertações e teses defendidas sobre Web Semântica e suas tecnologias. A integração entre comunidades científicas formadas em matrizes disciplinares diferentes encontra um ambiente favorável nos espaços de fronteira que demarcam as matrizes disciplinares. Comprova-se isso ao observar-se a comunidade pesquisadora sobre Web Semântica no Brasil, composta por pesquisadores de diversas matrizes como a Linguística, a Ciência da Informação e a Ciência da Computação.
Palavras-chave: Web Semântica. Cienciométrica. Comunidade científica.

Nível: Dissertação
Título: Web Semântica: aspectos interdisciplinares da gestão de recursos informacionais no âmbito da Ciência da Informação
Autor: RAMALHO, R. A. S.
Ano: 2006
Orientador: Silvana Aparecida Borsetti Gregorio Vidotti; Co-orientador: Mariângela Spotti Lopes Fujita
Linha de pesquisa: Informação e Tecnologia; Organização da Informação
Instituição: Universidade Estadual Paulista
Resumo: No âmbito da gestão de recursos informacionais os modelos e métodos de organização e recuperação de informações sempre estiveram condicionados às tecnologias utilizadas, de modo que com desenvolvimento e intensificação da utilização das tecnologias digitais uma nova gama de possibilidades vem sendo incorporada aos processos de produção, armazenamento, representação e recuperação de informações, atingindo um estágio em que os modelos clássicos de organização e recuperação de informações precisam ser (re)pensados sob diferentes perspectivas, pois os mesmos não parecem ser capazes de solucionar os problemas identificados no ambiente Web, evidenciando a necessidade de desenvolvimento de novas tecnologias que permitam otimizar a recuperação de informações em ambientes digitais. Nesse sentido, os estudos relacionados ao projeto Web Semântica vêm destacando-se como uma nova perspectiva no desenvolvimento de tecnologias que possibilitem um aumento na qualidade e relevância das informações recuperadas, a partir do desenvolvimento de instrumentos que permitam descrever formalmente, em um formato que possa ser processado por máquinas, os aspectos semânticos inerentes aos recursos informacionais, contribuindo para a identificação e contextualização das informações disponíveis no ambiente Web. Deste modo, a proposição deste trabalho é a realização de um estudo teórico e metodológico de caráter interdisciplinar acerca do projeto Web Semântica, buscando favorecer a “desmistificação” dos conceitos e tecnologias subjacentes e avaliar em que medida a área de Ciência da Informação pode contribuir para sua concretização, ressaltando os possíveis reflexos destas novas abordagens tecnológicas em seu corpus teórico. Assim, apresenta-se um levantamento bibliográfico acerca do desenvolvimento da Internet e os principais conceitos e tecnologias relacionados ao projeto Web Semântica, contextualizando-os a partir dos aportes teóricos da área de Ciência da Informação. Verificou-se que os conceitos e tecnologias subjacentes ao projeto Web Semântica podem ser considerados como uma renovação ou desdobramento dos métodos convencionais de representação, organização e recuperação de informações, apontando a possibilidade de contribuições da área de Ciência da Informação para o desenvolvimento do projeto Web Semântica, devido ao seu embasamento teórico referente a formas de representação e as práticas profissionais identificadas em seu campo de atuação, sendo evidente o caráter interdisciplinar que delinea o corpus teórico dos estudos relacionados ao projeto Web Semântica, apresentando-se como um

campo fértil para pesquisas e indicando a possibilidade de desenvolvimento de novos métodos de organização e recuperação de informações, construídos por meio de esforços interdisciplinares que favoreçam a junção da fundamentação teórica inerente à área de Ciência da Informação com as novas tecnologias emergentes da área de Ciência da Computação.

Palavras-chave: Web Semântica. Recuperação de informação. Ontologia. Gestão de recursos informacionais. Internet.

Nível: Dissertação
Título: Recursos tecno-metodológicos para a descrição e recuperação de informações na Web
Autor: SANTAREM SEGUNDO, J. E.
Ano: 2004
Orientador: Silvana Aparecida Borsetti Gregorio Vidotti
Linha de pesquisa: Informação e Tecnologia
Instituição: Universidade Estadual Paulista
Resumo: A tecnologia trouxe à Ciência da Informação uma nova partícula em seu objeto de estudo - a informação na Web; trouxe, também, uma aproximação muito grande entre as Ciências da Informação e da Computação. A Internet vem crescendo rapidamente, incrementando a explosão de informações, de formas a termos uma grande quantidade de informação disponível na Web. Desse modo, torna-se necessário investigar tecnologias para descrição e recuperação de informações que possibilitem a organização da informação digital no âmbito da World Wide Web. Valendo-se de pesquisa documental em fontes das áreas da Ciência da Computação e Ciência da Informação e da própria rede Internet foram analisadas as principais linguagens e os recursos para publicação de informações na Web, as formas de descrição e recuperação de informação, as propostas de novos padrões e de estrutura de dados e abordadas as novas ferramentas que vêm sendo discutidas e implementadas, objetivando a organização da informação digital. Verificou-se o delineamento de uma Web Semântica, que se trata de uma extensão da Web atual e que propõe uma nova arquitetura, de maneira que possamos dar significado a toda informação encontrada neste novo conceito de Internet. Tais aspectos permitem concluir que a criação da Web Semântica é uma questão de tempo e que, em breve, essa nova extensão da Web passará a ser um pedaço consistente e qualificado de informações dentro da Internet, possibilitando a várias comunidades a construção de conhecimento a partir de dados confiáveis encontrados na rede.
Palavras-chave: Internet. XML. Web Semântica. Linguagens de Marcação. Ferramentas de busca. Descrição e recuperação de informação na Web.

Apêndice C. Palavras-chave indicadas nas teses e dissertações analisadas

Ambiente de autoria	MarcOnt
Ambigüidade	Metadados
Arquitetura da informação	Método de análise de conteúdo
Arquitetura de metadados RDF	Modelo(s) de representação
BDTD	Modelos de recuperação de informação
Bibliotecas digitais	Ontologia(s)
Busca e recuperação de informação na Web	Organização da informação
Catologação automatizada	OWL
Ciência da Computação	Padrões de metadados
Ciência da Informação	Portais semânticos
Cienciometria	Portal corporativo
Comunidade científica	Produção de documento
Descrição de recursos na Internet	RDF
Descrição e recuperação de informação na web	Recuperação da(e) Informação
Dublin Core	Relações conceituais
Ferramentas de busca	Representação da Informação
Fontes heterogêneas	Representação do conhecimento
Gestão de Recursos Informacionais	Sistema legislativo
Gramática de valências	Taxonomia
IBICT	TEDE
Indexação	Teoria comunicativa da Terminologia
Indexação de imagens	Terminologia
Informação	Tesouro
Informação e tecnologia	Validação de ontologias
Informação legislativa	Web
Internet	Web Semântica
Interoperabilidade	
Linguagem de marcação XML/ XML	
Linguagem documentária	
Linguagens de marcação	
Linguagens de ontologias	
Log de eventos	
MARC21	